

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2005, 13:48:02 ; Search time 3265 Seconds
(without alignments)
10462.765 Million cell updates/sec

Title: US-10-810-788a-4
Perfect score: 705
Sequence: 1 aagcttcaacttgggaaa.....ctctgctcacttgaataacc 705

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:.*
1: gb ba:.*
2: gb htg:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_ste:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vi:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	653.8	92.7	84203	8	AC0005106	AC0005106 Genomic B
2	653.8	92.7	103223	8	AC007153	AC007153 Arabidops
3	73.4	10.4	177522	9	AC0093566	AC0093566 Homo sapi
4	71.8	10.2	174080	2	AF217246	AF217246 Homo sapi
5	71.6	10.2	186175	2	CR318664	CR318664 Danio rer
6	70.4	10.0	246589	2	CR450749	CR450749 Danio rer
7	70	9.9	156550	2	AC015830	AC015830 Homo sapi
8	68	9.6	281723	3	PF2929359	PF2929359 Plasmodiu
9	67.8	9.6	18556	9	AY527817	AY527817 Homo sapi
10	67.8	9.6	174741	9	AC040977	AC040977 Homo sapi
11	67.4	9.6	225328	2	CR847887	CR847887 Danio rer
12	67.2	9.5	165797	2	CR847851	CR847851 Danio rer
13	67	9.5	125623	3	AC115599	AC115599 Dictyoste
14	66.8	9.5	188394	2	CR753884	CR753884 Danio rer
15	66.6	9.4	156762	2	CR753796	CR753796 Danio rer
16	66.6	9.4	253755	2	CR812465	CR812465 Danio rer
17	66.2	9.4	209491	2	CR392365	CR392365 Danio rer
18	66	9.4	169448	2	CR385039	CR385039 Danio rer
19	66	9.4	273275	3	AE014828	AE014828 Plasmodiu

20	65.6	9.3	273275	3	AE014828	AE014828 Plasmodiu
21	65.4	9.3	177444	2	AC146461	AC146461 Pan trogl
22	65.4	9.3	181430	2	CR407555	CR407555 Danio rer
23	65.4	9.3	183344	9	AC147112	AC147112 Pan trogl
24	65	9.2	185342	2	CR387989	CR387989 Danio rer
25	64.8	9.2	149831	2	CR847943	CR847943 Danio rer
26	64.4	9.1	183648	3	AC117081	AC117081 Dictyoste
27	64.2	9.1	8056	6	AX559046	AX559046 Sequence
28	64.2	9.1	174574	2	CR759746	CR759746 Danio rer
29	64	9.1	191689	2	BX957320	BX957320 Danio rer
30	64	9.1	211643	5	BX530036	BX530036 Zebrafish
31	63.8	9.0	148486	2	CR753871	CR753871 Danio rer
32	63.8	9.0	333321	3	AC116986	AC116986 Dictyoste
33	63.6	9.0	250823	3	AE014821	AE014821 Plasmodiu
34	63.4	9.0	611	6	AX187051	AX187051 Sequence
35	63.2	9.0	253151	3	AE014842	AE014842 Plasmodiu
36	63	8.9	110000	2	PFMAL13_24	PFMAL13_24 o
37	63	8.9	146570	3	AC117076	AC117076 Dictyoste
38	63	8.9	182870	3	AC116960	AC116960 Dictyoste
39	62.8	8.9	250022	3	AE014824	AE014824 Plasmodiu
40	62.8	8.9	250793	2	CR388030	CR388030 Danio rer
41	62.6	8.9	1453	3	AF483206	AF483206 Plasmococ
42	62.6	8.9	8056	6	AX559046	AX559046 Sequence
43	62.6	8.9	85916	3	AC117080	AC117080 Dictyoste
44	62.6	8.9	113880	3	PFMAL3P4	PFMAL3P4 Plasmodiu
45	62.6	8.9	168069	2	CR812470	CR812470 Danio rer

ALIGNMENTS

RESULT 1	AC005106	84203 bp	DNA	linear	PLN 28-JUN-2000
LOCUS	AC005106	Genomic sequence for Arabidopsis thaliana BAC T25N20 from chromosome I, complete sequence.			
DEFINITION	AC005106	Genomic sequence for Arabidopsis thaliana BAC T25N20 from chromosome I, complete sequence.			
ACCESSION	AC005106	GI:6587719			
VERSION	AC005106.2				
KEYWORDS	HTG				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	1 (bases 1 to 84203)				
AUTHORS	Johnson-Hopson, C., Brooks, S., Buehler, E., Chao, Q., Khan, S., Kim, C., Shinn, P., Altati, H., Bei, Q., Chin, C., Chlou, J., Choi, E., Conn, L., Conway, A., Gonzales, A., Hansen, N., Howling, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharasy, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thavert, A., Tortum, M., Vayberg, M., Yu, G., Federspiel, N.A., Theologis, A. and Ecker, J.R.				
TITLE	Genomic sequence for Arabidopsis thaliana BAC T25N20 from chromosome I				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 84203)				
AUTHORS	Ecker, J.R.				
TITLE	Direct Submision				
JOURNAL	Submitted (16-JUN-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	3 (bases 1 to 84203)				
AUTHORS	Ecker, J.R.				
TITLE	Direct Submision				
JOURNAL	Submitted (31-JUL-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	4 (bases 1 to 84203)				
AUTHORS	Ecker, J.R.				
TITLE	Direct Submision				
JOURNAL	Submitted (01-DEC-1998) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				

REFERENCE AUTHORS	5 (bases 1 to 8403) Shim,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altati,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howling,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mkharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA 6 (bases 1 to 84203)
REFERENCE AUTHORS	Ecker,J.R.
TITLE JOURNAL	Direct Submission Submitted (10-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA 7 (bases 1 to 84203)
REFERENCE AUTHORS	Chen,K., Shim,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altati,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howling,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mkharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA On Dec 16, 1999 this sequence version replaced gi:3366536.
COMMENT FEATURES source	Location/Qualifiers 1. 84203 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="1" /clone="T25N20" 272. 454 /note="hypothetical protein" /codon_start=1 /evidence=not experimental /product="T25N20.1" /protein_id="AAE79720.1" /db_xref="GI:8778712" /translation="MNPSTYVLSLQKFCINLEAMATIRPHITHFHAHKIPPTKINSIKRYANNKCYIIM" complement(join(1491..2159,2389..2452,2568..2697,3061..3118,3369..3457,4022..4139,4246..4353)) /note="similar to YUP8H12.1 gb AAE71441.1; similar to EST gb A1996871.1" /codon_start=1 /evidence=not experimental /product="T25N20.2" /protein_id="AAE79741.1" /db_xref="GI:8778713" /translation="MGKKEKDHSVSDDKVEAVLHLHLKSHSLTLQKFCNRACVGAFILKGNVYKAKADKSCISWRSS-IGISLADSEFTLAEGLAYVGLDDEKCPVLAITMSRNVEQFVILFDATVAVLITLITLPSRLNVALITLCSNTALPPPEPFIYKTPKLIQISIFHEHTCYTQNSRGGITFPDLSITAINIYSMODFQDSFPYDPSSTYSPRVSILRDTSLKSDTKIGSCASSRPAFTVSRDGLDYKPMCLTLTIDTSSTKLGNTAFVYISPLNARSFPASPARSEPGGRRSPFASTPMPATTDHSHIGTLRDLRNPSPFQPAIFFRRESHVSKSEKPRDSFVQFLKFYRRPYDEWITYSKMRPLGGLVSHVIOIRRHHSLSQRP"
CDS	join(9304..11268,11387..11513,11607..11692,12103..12341,12698..12784,13257..13356,13451..14263) /note="unknown protein; similar to ESTs gb AWS60017.1, and gb A1998699.1" /codon_start=1 /evidence=not experimental /product="T25N20.3" /protein_id="AAE79721.1" /db_xref="GI:8778713"
REFERENCE AUTHORS	5 (bases 1 to 8403) Shim,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altati,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howling,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mkharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA 6 (bases 1 to 84203)
REFERENCE AUTHORS	Ecker,J.R.
TITLE JOURNAL	Direct Submission Submitted (10-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA 7 (bases 1 to 84203)
REFERENCE AUTHORS	Chen,K., Shim,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altati,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howling,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mkharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA On Dec 16, 1999 this sequence version replaced gi:3366536.
COMMENT FEATURES source	Location/Qualifiers 1. 84203 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="1" /clone="T25N20" 272. 454 /note="hypothetical protein" /codon_start=1 /evidence=not experimental /product="T25N20.1" /protein_id="AAE79720.1" /db_xref="GI:8778712" /translation="MNPSTYVLSLQKFCINLEAMATIRPHITHFHAHKIPPTKINSIKRYANNKCYIIM" complement(join(1491..2159,2389..2452,2568..2697,3061..3118,3369..3457,4022..4139,4246..4353)) /note="similar to YUP8H12.1 gb AAE71441.1; similar to EST gb A1996871.1" /codon_start=1 /evidence=not experimental /product="T25N20.2" /protein_id="AAE79741.1" /db_xref="GI:8778713" /translation="MGKKEKDHSVSDDKVEAVLHLHLKSHSLTLQKFCNRACVGAFILKGNVYKAKADKSCISWRSS-IGISLADSEFTLAEGLAYVGLDDEKCPVLAITMSRNVEQFVILFDATVAVLITLITLPSRLNVALITLCSNTALPPPEPFIYKTPKLIQISIFHEHTCYTQNSRGGITFPDLSITAINIYSMODFQDSFPYDPSSTYSPRVSILRDTSLKSDTKIGSCASSRPAFTVSRDGLDYKPMCLTLTIDTSSTKLGNTAFVYISPLNARSFPASPARSEPGGRRSPFASTPMPATTDHSHIGTLRDLRNPSPFQPAIFFRRESHVSKSEKPRDSFVQFLKFYRRPYDEWITYSKMRPLGGLVSHVIOIRRHHSLSQRP"
CDS	join(9304..11268,11387..11513,11607..11692,12103..12341,12698..12784,13257..13356,13451..14263) /note="unknown protein; similar to ESTs gb AWS60017.1, and gb A1998699.1" /codon_start=1 /evidence=not experimental /product="T25N20.3" /protein_id="AAE79721.1" /db_xref="GI:8778713"
REFERENCE AUTHORS	5 (bases 1 to 8403) Shim,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altati,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howling,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mkharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (16-DEC-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA 6 (bases 1 to 84203)
REFERENCE AUTHORS	Ecker,J.R.
TITLE JOURNAL	Direct Submission Submitted (10-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA 7 (bases 1 to 84203)
REFERENCE AUTHORS	Chen,K., Shim,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altati,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howling,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mkharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE JOURNAL	Direct Submission Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA On Dec 16, 1999 this sequence version replaced gi:3366536.
COMMENT FEATURES source	Location/Qualifiers 1. 84203 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="1" /clone="T25N20" 272. 454 /note="hypothetical protein" /codon_start=1 /evidence=not experimental /product="T25N20.1" /protein_id="AAE79720.1" /db_xref="GI:8778712" /translation="MNPSTYVLSLQKFCINLEAMATIRPHITHFHAHKIPPTKINSIKRYANNKCYIIM" complement(join(1491..2159,2389..2452,2568..2697,3061..3118,3369..3457,4022..4139,4246..4353)) /note="similar to YUP8H12.1 gb AAE71441.1; similar to EST gb A1996871.1" /codon_start=1 /evidence=not experimental /product="T25N20.2" /protein_id="AAE79741.1" /db_xref="GI:8778713" /translation="MGKKEKD

[illegible]

```

ACCESSION      AC007153
VERSION        AC007153.2  GI:4580365
KEYWORDS       HTG.
SOURCE         Arabidopsis thaliana (thale cress)
ORGANISM       Arabidopsis thaliana
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
                rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS        Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
                Alfati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
                Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
                Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vydrotskaia,V.,
                Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
JOURNAL         Submitted (24-MAR-1999) DNA Sequencing and Technology Center,
                Stanford University, 855 California Avenue, Palo Alto, CA 94304,
                USA
TITLE          Direct Submission
                Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
                Stanford University, 855 California Avenue, Palo Alto, CA 94304,
                USA
REFERENCE
AUTHORS        Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
                Alfati,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
                Gonzalez,A., Kremetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
                Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vydrotskaia,V.,
                Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
JOURNAL         Submitted (08-APR-1999) DNA Sequencing and Technology Center,
                Stanford University, 855 California Avenue, Palo Alto, CA 94304,
                USA
TITLE          Direct Submission
                Submitted (17-MAY-1999) DNA Sequencing and Technology Center,
                Stanford University, 855 California Avenue, Palo Alto, CA 94304,
                USA
COMMENT
                On Apr 9, 1999 this sequence version replaced gi:4580804.
                Bases 1-33654 of clone of clone F3F20 overlap with bases
                50550-84203 of 'TAMU' clone T25N20, gb|AC005106.
                e-mail for correspondence: araba@sequence.stanford.edu
                Genes with similarity to proteins in the databases are described as
                'putative', '-like' or 'similar to'. Genes that have EST
                similarity but no significant protein similarity are described as
                'unknown proteins'. Genes that are annotated based only on gene
                prediction software are described as 'hypothetical proteins'.
                The software programs used to predict genes include: Grail
                (Informatics Group, Oak Ridge National Laboratory,
                http://comprobi.cornell.gov/section/index.html), GENSCAN (Chris Burge,
                http://genomic.stanford.edu/~chris/GENSCANW.html), Fexa (V.Solovyev
                & A.Salamov, Sanger Centre, http://genomic.sanger.ac.uk/), and
                NecPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of
                Denmark, http://www.cbs.dtu.dk/NecPlantGene.html).
FEATURES
SOURCE
                1..103223
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /culivar="Columbia"
                /db_xref="caxon:3702"
                /chromosome="1"
                /clone="F3F20"
                1..33654
                /note="overlap with bases 50550-84203 of 'TAMU' clone
                T25N20, gb|AC005106 Features from bases 1 to 29,000 are
                annotated on overlapping clone T25N20, gb|AC005106."
misc_feature

```

```
gene
CDS
    complement(29404..40201)
    /gene="F3F20.1"
    /note="Highly similar to putative callose synthase catalytic subunit; Highly similar to cotton putative callose synthase catalytic subunit, gb|A025952"
    /codon_start=1
    /protein_id="AAD30609.1"
    /db_xref="GI:4836907"
    /translation="MAQREPPPPPPRRILRTQTVSGSLGEAMLDSEVVPSSIVEIAP
    IIRVANEVASNPVAYLCRFYAFKRAHLDPPSGRQOFTALLOLERENETTL
    AROKSDAREMOSFOHYKKYKYOIALNLADKADRAOLTKRAYOTALVFEVKAQVOT
    EBEVADDEVNLIYDIDLTQIYVPPYNIPLDPSNOAMLPETIOAAVAARNG
    LPTAGHKKKLDLIDMLQSMFGOKVNLNREHLILLANVHAKOPKPDQOKL
    DRAALTYVKKLFRNYKKWKCYLGRSSLMPLTIOQVQOKLLYGLLWGEAN
    LAFMPCLCYIYHHLGMLAGSVPMTGSHVKAQGEDEAFQKVTVPYOTISKA
    KSGRQKSHSVWRNYDDLNIEFWSIRCFRLGPMRADPFCOTAELELSKES
    NSGDRMGKVNFEIRSPWNIHPSRDLMSFYILCIOMIVITANNSGSLAPFOODV
    FKLVSVPITAAIKLAQAVLDIALSWKARHSLSLYKARYWKVGAAYVYVMAVT
    VAYSKNASGFSOTIKMFGHSHNSPLIYAILLYSPNMLSLPLPEPRKRLLE
    RSDYKIMLMWMMWSQSPFSSMOQIPLVGPTDINRIHIVYSWHEFPKAPKINVL
    VIALVSPVLIVYEMDQIYVAIVSTLVGLNGAFRLGEIRLGLMKSFPQSLPGAFN
    DCLVDPNSDDTCKKFRATFSRKPDPSSKQKEARFAOMNKLISFREDLID
    REMELLIVYMSDPLDIRMPRLIASIPIALDMAKDSNGKRELKRIKIVDSWT
    CAVRECYAFKQILNYLVGERGOYINDIFSKIDHIKETITILNLSALPDLYQ
    FRLIYELNREEDKOQIVILNMLELITRDMEEVPSANISVNFOSPFLKXL
    GKKKQIKRLHLITVYESAMDPVSNLEARRLTFPENSLSFMDPPAPKIRMLSEFVL
    TTYFSEDFVFSIFGLEQNEADVSLIFLOKIFPDEMTNLFEMDKGCEBESILRARLD
    BEELRLMASYRGOTLTQVGRMYKRALELQAFDMAKDELLKGYKLELISEAS
    KSGSLMAOCCALADMKPTFVVSQOQYSHKSGDORADILRLMTTYSIVAYIDE
    VQOTKESYKGTBEKIYYSALYKAAPOTKPMSSSEVQTLDOIYIKLPGRIIDEG
    KENONHAIIFTRBGLQITIDMKNQYMEAFKRWLDQEFKSHGVCPTYLKARE
    H1PTGVSLSLAMPNQSNSFVITIGORVLASPLKXVPHYGHPIIDRLPHLRGPNST
    LRBGVYTHIEYIYQVKGKRDVGLQIISMPAKIANGSGEOTLSDLRYLRLEDFEML
    SCFYITIGFYFTMLTVLVYVLYLGRVILVLSGLEEGSSORAFNNKPLEALASO
    SFVOIGFIMALPMMEIGLERGHNAIIEFVLMQOLASVFFPOIGTTHYGRTLF
    HGGAERYGTGRGVYVYHAKFAENRYRYSRSHYKGIELMILLIYOTIFQOSIRGVTY
    ILITVSIWVAVTWLFAPLFNPSGFEWQKIVDDWDMKNKILYNRGQISVPPKESVI
    LBELELEHLSHSGVRGILTLEIFLALRPIFOYGLVYHLSIFKGNQSFVYGAISWVI
    LEILLIVKLGVRGRRTSTNFOLEFRIIKGLVLEFVALIFLALPILITIKDLPICM
    LAEPFGMGLMLIAOCKPLIOLQIGIOWSSSVTLARGYEIVMGILLFTPYAPLAMPFV
    SFQOTRMLFNQASRGLQISRLIGGORKRSKXKXK"
    complement(44167..45812)
    /gene="F3F20.2"
    /note="Hypoetical protein"
    /codon_start=1
    /protein_id="AAD30611.1"
    /db_xref="GI:4836909"
    /translation="MESNGGGEVRRVNLVYFLSRSGVHDHPLILKVLHLSNGVFLR
    DYKMLADARGMDPAFSMSCKRYKYNQVWODLDDLLIPISNREVLAGSESLI
    SSPEKYVNEKAMWYRNGCIDAEKLOKLTYSKTIKQESVSTSTSTSTSTSTSTST
    BESTTNEGFVLKKQDPKTVSGGSDSTENGSDNDESGRPSVSTSSSTKNSY
    SSVRSASHVRLNLMKCGGLDINDAVLVLSRSGAGPMEBERARYOYQONARASF
    BEAWSGIRKMETIEFCCKPVASKPSMALCSGCLPEPEKMSHMKCRGKNSA
    NNDLMTSNVYVPROQRNIPGNPLIGHORVLTTLTKE"
    /gene="F3F20.3"
    complement(47118..50318)
    /gene="F3F20.3"
    complement(5019(47118)..48467,48761..49149,49186..49302,
    49523..49979,50064..50318))
```

```

    /gene="F3F20.3"
    /note="similar to Na/H antiporter proteins; N-terminal half of protein is similar to Na/H antiporter proteins"
    /codon_start=1
    /protein_id="AAD30610.1"
    /db_xref="GI:4836908"
    /translation="SSGAPLAVTNPNYIIEBSRREKICYDOSLLFEKRGKMSG
    STLASLPEFTIQLFVANSYRVLVLPVLPVPAOILCGLSPSVLQGTRETI
    AHVFPFRFMVJETEPANLALVYNIPLFGGMDLRMRITELKXVILVGLVALPVG
    AFVLYPFGHSDPKIISGCVFMSVALACNMPDLALDLIDKSLDMKRTALCAIV
    TDLCTWVLVPGASFSKSGTNWKNMPFITAITAYLICIPIRGGIMVIFKYTKA
    GGVLICGLTTCGCHSITNGAPLFGSLTIDHDIRMEIEKILDELSTGLMLFYIC
    GLRADIGMLQTDKRMVYVVCSSFLVAVITVTSLSLMLPMDAPRIGALMTKG
    TISLVVNLNAGRDTKALDSPYTHMTLALVMSLVPEPLAFAPKPKLAAHYKRTVQ
    KIKGETELRVLCVHLPMVSGITNLIQVSNATKQSPLSVFALHVELGRTASLLI
    MDCKEKPANFSDRYAEDQIAETEAEMVNNDAMTVOITDVALSPYATMHEDIIVLA
    EKRVCFTILPRYHKLTPDGRGEGNSHAEINONVLSHAPCSVGLVVRGMAVRS
    SPRGSMREVMAMLPVGGPDRDEALSAYRMCOHYIKTVVPRGREALISGCVLA
    AERYEKQYDDECITIEFNKTYNDSSVKTIKRYNDGQOTIATIREMEDNSISGLVY
    GRGVNSDSPVTAGLNDWSSPELGTIGDTLASNFTWASVLVIOQYSATKROAAVTA
    AATTVMGAVAGVTCGNLLESAGGDAMKTRDAHEPFKMSYEDDEDEBHOYIHR"
    51395..53210
    /gene="F3F20.4"
    /note="Similar to hexosaminidase; Similar to hexosaminidase, gi|1170249, gi|1076942 and others"
    /codon_start=1
    /protein_id="AAD30612.1"
    /db_xref="GI:4836910"
    /translation="MLTSLKRFVILIPILFTTLSPSLPISALPINIKRPRFLSPQ
    HAAILSPNFTLAPHEOYLSVYRKNVLIISENSPLISYPVKLMKRYTLRNLTVT
    VYDFSLPLHNGDESKYSIPIGFSFAHLIASAVAMKMLETFSQIMGTSDCLP
    VYIDVDSPLDGFGRVILDTISRNRYGVDDIMETIKAMSKANLNFPMHITDSQFLV
    LSEBPLAKSGSGPMVYTPPDVSKIVOYGEGRVYRPEIDTPSSWGAAYEI
    VYCANFPMWPAKSMWEERLASRPGQNLPLSKPYKEVYVNIQDIYNOPPSRFFRG
    GBEVIFGCKRIDPAINSLSSGTSLSQLEKINSTLPIYIVSONRTVYVMEVDLDAQ
    IKADSPVLPKEHTIQLTNNGENPRTYVIAAVRYVSSSEYVYIDCGGFGKNDIS
    IYQKSGSGSCWCAFPKTMQSIYNYDIADLDLNEERKVLGGEVALLMSQASVTVDS
    RLMPSRVALAESLSCNRDERGVKRCGEAVRLNIRYVYKRGIGAEPIQLMCLKN
    PGMCTVHALADQ"
    53685..55189
    /gene="F3F20.5"
    53685..55199
    /gene="F3F20.5"
    /note="Hypoetical protein"
    /codon_start=1
    /protein_id="AAD30626.1"
    /db_xref="GI:4836924"
    /translation="MSVVRMPRVLPSSLQILKKQKNPTALKLFBAKERPPSYGH
    NSVAVATMIDIIGKSNRVLEMKYIREMKEDSCCKDSVFAVIRTFPSRGRLEDAIS
    Query Match 92.7%; Score 653.8; DB 8; Length 103223;
    Best Local Similarity 99.1%; Pred. No. 6.7e-94;
    Matches 699; Conservative 0; Mismatches 2; Indels 4; Gaps 4;
    QY 1 AAGCTTTAAATTTGGGAAAAATTTAATATATGATTAATATG-CAATATAAAGTAG 59
    DB 20166 AAGCTTTAAATTTGGGAAAAATTTAATATATGATTAATATGCAATATAAAGTAG 20225
    QY 60 ATGAGAGTTTTTTTAAATTTTATTAATAAATAA-TTTTTTGCATGAATGTTTTTA 118
    DB 20226 ATGAGAGTTTTTTTAAATTTTATTAATAAATAAATTTTTTGCATGAATGTTTTTA 20285
    QY 119 GATTAATTTTGAATTAACCAAGTATTTATTTCCGACGAATTGAGTCTTACGAGGAG 178
    DB 20286 GATTAATTTTGAATTAACCAATATATTTTCCGACGAATTGAGTCTTACGAGGAG 20345
    QY 179 TCTGTCTTAAACCTGTAAGTTTGGCTTACCAACCAATGACCAATGCAAAAGGTAA 238
    DB 20346 TCTGTCTTAAAC-CTCGTAAGTTTGGCTTACCAACCAATGACCAATGCAAAAGGTAA 20404
    QY 239 CGACCATACCGGTATAATATTCTTAACCGGTTTATAGTTTACATTAATCATTTACTAA 298
```



```

Db      20405 CGA-CATTAACGGTTAAATTAATTCATACCCGTTTATTAAGTTTACATTAATCATTTACTAA 20463
Qy      299   TCGGCGGTAAATTAAGTTTAATGAATNGGTTATTTTGTAGGACAGTGAATTTATTA 358
Db      20464 TCCGCGGTAAATTAAGTTTATGAATNGGTTATTTTGTAGGACAGTGAATTTATTA 20523
Qy      359   ATTTTGTAGTACTGTTTTCCTTTTGGGTTCAACTAGTTACTTTTTCCTTGTGACATC 418
Db      20524 ATTTTGTAGTACTGTTTTCCTTTTGGGTTCAACTAGTTACTTTTTCCTTGTGACATC 20583
Qy      419   AAAATTATTTGTAGACAGATGCTCCATATATAGATGCGTAATGAAATGAAATTTAGTAA 478
Db      20584 AAAATTATTTGTAGACAGATGCTCCATATATAGATGCGTAATGAAATGAAATTTAGTAA 20643
Qy      479   TAAATTAATTAATTAAGTGAACAAAAAATTAAGTGAACAAAAAGTAACTGAAAGAAAG 538
Db      20644 TAAATTAATTAATTAAGTGAACAAAAAATTAAGTGAACAAAAAGTAACTGAAAGAAAG 20703
Qy      539   CTATTATGCTTGCGCAAGATTCATTAAGTTATTTTACCAAAAGAAAGAGTACTTGCC 598
Db      20704 CTATTATGCTTGCGCAAGATTCATTAAGTTATTTTACCAAAAGAAAGAGTACTTGCC 20763
Qy      599   TTGCTCTCTCTCTCTTTTAAATGTTAACTGTAAGATGAGAGACTCAATTTCTAGTT 658
Db      20764 TTGCTCTCTCTCTCTTTTAAATGTTAACTGTAAGATGAGAGACTCAATTTCTAGTT 20823
Qy      659   CATCAAAACATATTTGGCGTTATTTTCTGTGCTCCTTGAAATA 703
Db      20824 CATCAAAACATATTTGGCGTTATTTTCTGTGCTCCTTGAAATA 20868

```

```

RESULT 3
AC093566 177522 bp DNA linear PRI 29-NOV-2001
LOCUS Homo sapiens chromosome 8, clone CTD-2017M22, complete sequence.
AC093566
VERSION AC093566.3 GI:16905291
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 177522)
AUTHORS Birren, B., Linton, L., Nuebaum, C. and Lander, E.
TITLE Homo sapiens chromosome 8, clone CTD-2017M22
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177522)
AUTHORS Birren, B., Linton, L., Nuebaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chararo, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meidrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trifillio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

```

TITLE
JOURNAL
Submitted (03-SEP-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

REFERENCE 3 (bases 1 to 177522)
AUTHORS Birren, B., Linton, L., Nuebaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chararo, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meidrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trifillio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 13, 2001 this sequence version replaced gi:16041411.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12730
Center clone name: 2017_M_22
-----
FEATURES
source
Location/Qualifiers
1. 177522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="CTD-2017M22"
/clone_1lb="CITD Human BAC"
106..781
repeat_region
/rpt_family="MER21B"
794..1024
repeat_region
/rpt_family="MIR"
1654..1695
repeat_region
/rpt_family="L2"
2458..3178
repeat_region
/rpt_family="L1MA7"
3180..3323
repeat_region
/rpt_family="AT-rich"
3323..3986
repeat_region
/rpt_family="L1MCD"
3986..4520
repeat_region
/rpt_family="L1MA3"
4521..5565
repeat_region
/rpt_family="THB1-int"
5565..7819
repeat_region
/complement(5560..7819)
/rpt_family="L1M1"
7819..8206
repeat_region
/complement(7814..8206)
/rpt_family="L1MCD"
8206..8511
repeat_region
/complement(8207..8511)
/rpt_family="AluSx"
8511..8628
repeat_region
/complement(8512..8628)
/rpt_family="L1MCD"
8629..8683
repeat_region

```

[illegible]


```

QY      306 GTAATTAGTTTATGAAATGCTGTTATTTTGTAGTCACGTGAAATTTAATTTTTT 365
Db      92069 TTAATAATATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 92010
QY      366 AGTACTGTTTCTTTCTTTTGGTTCACAGTACTGTTTCTTCTGATCAATCAATTA 425
Db      92009 TTTAAATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTT 91950
QY      426 TTGTAGACGAGTGTCATATATATAGTGTGAAATGAAATGAAATATTTGATTAATAA 485
Db      91949 TATATATATATATATATATATATATATATATATATATATATATATATTA 91890
QY      486 ATATAGAAAGTGACCAAAAAAATTTAGTGAAAAAGTAACTGAAAAGAAAGCTTAT 543
Db      91889 ATTAATAATATATATATATATATATATATATATATATATATATATATAT 91832

RESULT 6
CR450749 246589 bp DNA linear HTG 24-MAY-2004
LOCUS     Danio rerio clone DKEY-206, *** SEQUENCING IN PROGRESS ***, 11
DEFINITION
unordered pieces.
ACCESSION CR450749.1 GI:47604407
VERSION   HTG; HTGS_PHASE1.
KEYWORDS  Danio rerio (zebrafish)
SOURCE    Danio rerio
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 246589)
Sim8.S.
Direct Submission
Submitted (21-MAY-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZK2J6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 243112 bases at least Q40
Consensus quality: 243477 bases at least Q30
Consensus quality: 243996 bases at least Q20
Insert size: 245589; sum-of-contigs
Insert size: 241608; 2.3% error; agarose-fp
Quality coverage: 7.87x in Q20 bases; sum-of-contigs Quality
coverage: 8.07x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 28228: contig of 28228 bp in length
* 28229 28328: gap of 100 bp
* 28329 39447: contig of 1119 bp in length
* 39448 39547: gap of 100 bp
* 39548 48874: contig of 9327 bp in length
* 48875 48974: gap of 100 bp
* 48975 97551: contig of 4857 bp in length
* 97552 97651: gap of 100 bp
* 97652 184544: contig of 86893 bp in length
* 184545 184644: gap of 100 bp

```

```

* 184645 194757: contig of 10113 bp in length
* 194758 194857: gap of 100 bp
* 194858 209127: contig of 14270 bp in length
* 209128 209227: gap of 100 bp
* 209228 224943: contig of 15716 bp in length
* 224944 225043: gap of 100 bp
* 225044 230389: contig of 5346 bp in length
* 230390 230490: gap of 100 bp
* 230490 236916: contig of 6427 bp in length
* 236917 237016: gap of 100 bp
* 237017 246589: contig of 9573 bp in length.

FEATURES
    source
        1..246589
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="DKEY-206"
            /clone_1lb="DanioKey"
            1..28228
                /note="assembly fragment:01036
                fragment_chain:1"
                28329..39447
                    /note="assembly fragment:00514
                    fragment_chain:1"
                    39548..48874
                        /note="assembly fragment:00184
                        fragment_chain:1"
                        48975..97551
                            /note="assembly fragment:01406
                            fragment_chain:1"
                            97652..184544
                                /note="assembly fragment:02042
                                fragment_chain:1"
                                184645..194757
                                    /note="assembly fragment:00397
                                    fragment_chain:1"
                                    194858..209127
                                        /note="assembly fragment:00656
                                        fragment_chain:1"
                                        209228..224943
                                            /note="assembly fragment:00820
                                            fragment_chain:1"
                                            225044..230389
                                                /note="assembly fragment:00033
                                                fragment_chain:1"
                                                230490..236916
                                                    /note="assembly fragment:00101
                                                    fragment_chain:1"
                                                    237017..246589
                                                        /note="assembly fragment:00288.0"

ORIGIN
Query Match      10.0%; Score 70.4; DB 2; Length 246589;
Best Local Similarity 32.9%; Pred. No. 0.019;
Matches 188; Conservative 0; Mismatches 383; Indels 0; Gaps 0;

QY      2 AGCTTTAAATTTGGGAAATTTATATATATATATATATATATATATATATATAT 61
Db      184659 AANTNNNTTATTTNTTANTTANAATTTATTTTNTNNNNNANNTNNNNNTNNANNNNN 184718
QY      62 GAGAGTTTTTTTAAATTTTATATATATATATATATATATATATATATATATATAT 121
Db      184719 ANNNNTTTAAANNAANTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 184778
QY      122 AAAATTTTGACATTAACCAAGTATATTTTCCGCCACGAATTGAGTCTACGAGAGATGCC 181
Db      184779 NNNNNNNNNNNNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 184838
QY      182 TGCTTTAACTGCTAAAGTTTGCTTACCCACCAATCCCAATCAAGTAAAGCA 241
Db      184839 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 184898
QY      242 CCATACCGGTTAATATATCTTAACCGGTTATATAGTTTACATAATCATTTACTAATCC 301

```

[illegible]

* and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1	901: contig of 901 bp in length
902	1001: gap of 100 bp
1002	1895: contig of 894 bp in length
1896	1995: gap of 100 bp
1996	2901: contig of 906 bp in length
2902	3001: gap of 100 bp
3002	3906: contig of 905 bp in length
3907	4006: gap of 100 bp
4007	4924: contig of 918 bp in length
4925	5024: gap of 100 bp
5025	5945: contig of 921 bp in length
5946	6045: gap of 100 bp
6046	6633: contig of 918 bp in length
6634	7063: gap of 100 bp
7064	7999: contig of 936 bp in length
8000	8099: gap of 100 bp
8100	9123: contig of 924 bp in length
9024	9123: gap of 100 bp
9124	9975: contig of 852 bp in length
9976	10075: gap of 100 bp
10076	10999: contig of 924 bp in length
11000	11099: gap of 100 bp
11100	11997: contig of 898 bp in length
11998	12097: gap of 100 bp
12098	12998: contig of 901 bp in length
12999	13098: gap of 100 bp
13099	14001: contig of 903 bp in length
14002	14101: gap of 100 bp
14102	15033: contig of 932 bp in length
15034	15133: gap of 100 bp
15134	16056: contig of 923 bp in length
16057	16156: gap of 100 bp
16157	17082: contig of 926 bp in length
17083	17182: gap of 100 bp
17183	18104: contig of 922 bp in length
18105	18204: gap of 100 bp
18205	19081: contig of 877 bp in length
19082	19181: gap of 100 bp
19182	20148: contig of 967 bp in length
20149	20248: gap of 100 bp
20249	21156: contig of 908 bp in length
21157	21256: gap of 100 bp
21257	22141: contig of 885 bp in length
22142	22241: gap of 100 bp
22242	23177: contig of 936 bp in length
23178	23277: gap of 100 bp
23278	24186: contig of 919 bp in length
24187	24296: gap of 100 bp
24297	25254: contig of 928 bp in length
25255	25344: gap of 100 bp
25345	26250: contig of 926 bp in length
26251	26350: gap of 100 bp
26350	27292: contig of 942 bp in length
27292	27392: gap of 100 bp
27392	28324: contig of 932 bp in length
28325	28424: gap of 100 bp
28425	29358: contig of 934 bp in length
29359	29458: gap of 100 bp
29459	30340: contig of 882 bp in length
30341	30440: gap of 100 bp
30441	31139: contig of 899 bp in length
31139	31439: gap of 100 bp
31440	32360: contig of 921 bp in length
32361	32460: gap of 100 bp
32461	33339: contig of 879 bp in length

JOURNAL	Nature 419 (6906), 527-531 (2002)
MEDLINE	22255708
PUBMED	12368867
REFERENCE	2 (bases 1 to 281723)
AUTHORS	Sanders,M., Hauser,H., Baker,S., Unwin,L., Mungall,K., Berriaman,M., Pain,A., Hall,N., Bowman,S., Churcher,C., Quail,M. and Barrett,B.
TITLE	Direct Submission (20-SEP-2002) P.falciparum Genome Sequencing Consortium, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
JOURNAL	For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum.
COMMENT	Location/Qualifiers 1..281723
FEATURES	/organism="Plasmodium falciparum 3D7" /mol_type="genomic DNA" /isolate="J3D7" /db_xref="taxon:36329" /chromosome="9" 1017..1934 /gene="PF11520w" 1017..1934 /gene="PF11520w" /codon_start=1 /product="hypothetical protein" /protein_id="CAD51990.1" /db_xref="GI:23505210" /db_xref="UniProt/TREMBL:Q812K2" /translation="MLSIASTFGSYFSSCSVDYLNNEANKEPSSDEKINFRKRNVSSSKDKTININDHDKILNESDSHDINKINDLEAYATONNNKNKIYEHHIGINNEKENIDHQEVOISKDSHONGNEHQNMKSNETTLDEHKSEENALPHDHKEGAMNEKPNDIHSENAFLPYDMKLDDKEYPNPIKSHEQDCSDKSFDBCARKEELNHNKKKVNYIYNENAESNNYSFPNEKKYRLIGFSRMSGVITINEKKGGIANNCTVLVDDLFPSVSNYLSKMPDYLTNPKEEINKNE" join(3337..3354,3499..3577,3792..4561) /gene="PF11525w" join(3337..3354,3499..3577,3792..4561) /note="Moderate similarity to Mus musculus unknown TR:AAH06701 (EMBL:BC06701) (297 aa) fasta scores: E(): 0.028, 28.962% id in 183 aa. ProfileScan hit to PS50030, Ubiquitin-associated domain." /codon_start=1 /product="hypothetical protein" /protein_id="CAD51991.1" /db_xref="GI:23505211" /db_xref="UniProt/TREMBL:Q812K1" /translation="MDSENKLVQLVPMGSKSRISOKVIOKGAKTIIDAISWIELLDFTENNTBELADVNVTKDSEKGSFKNKSLKSEHNDENNITSCCEKSKSPERAKKALELOCKIREKRLKEKEBELEKKEKRIAMTKEMQRKEQLSEYERKKYTESLERAKEHKEKEKELIKREYAEKFGIEYKIKLODLTENEKREDIALIFNNLKONYKDTKKOELLASINITRTYFNSNYDLLEKKYOKIKEENKIFVEKIKIFEEMLIFFLVGFEDYGRMKAFIN"
gene	complement(join(4741..4779,5025..5294,5425..5555,5783..6005)) /gene="PF11530c" 5783..6005) complement(join(4741..4779,5025..5294,5425..5555,5783..6005)) /gene="PF11530c" /note="Signal anchor predicted for PF11530c by SignalP 2.0 HMM (Signal peptide probability 0.005, signal anchor probability 0.937) with cleavage site probability 0.004 between residues 28 and 29 1 probable transmembrane helix predicted for PF11530c by TMHMM2.0 at aa 12-31"
CDS	/codon_start=1 /product="hypothetical protein" /protein_id="CAD51992.1" /db_xref="GI:23505212" /db_xref="UniProt/TREMBL:Q812K0" /translation="NNIIYLINTDYTYLPFYFPLIICWTWYIQVVLKLDONKXSTEINNEKINHCKQLDKNNNSQLENLYKLIKNTKDNVEDLLSRKTASYIVLRALFLIHNKFIHLIRKNCIPFIRNNITDIRNLKTLTKVKVIQFKQDMNICLSAENVSPKNNK
gene	join(8267..8365,8749..9405) /gene="PF11535w" join(8267..8365,8749..9405) /gene="PF11535w" /codon_start=1 /product="hypothetical protein" /protein_id="CAD51993.1" /db_xref="GI:23505213" /db_xref="UniProt/TREMBL:Q812J9" /translation="MEKTTEDDFIDSTEETLKKKITPTDLYTEDELPEIPISGLSKRGVLSKEDYSTHYQKKRNERNKIIEEVNNIITYDINDLNKSGFFPINATHLHFQOEKISTILLPIRYISKKSEISSYIDINNKAIAEKIKRKYYASKOLSYNNHNNVCNNSENEYEIIINKDIGLPGHISNTFYILNSOKEIKIKLSDNIYKONIIMNSNPRBERINSEDIKKKEIKDPNYLOCIIYTTHL"
CDS	join(10145..10388,10425..10511,10701..12151,12317..12646) /gene="PF11540w" join(10145..10388,10425..10511,10701..12151,12317..12646) /gene="PF11540w" /codon_start=1 /product="hypothetical protein" /protein_id="CAD51994.1" /db_xref="GI:23505214" /db_xref="UniProt/TREMBL:Q812J8" /translation="MEDIKSYEEKLKHARQHEPRLPEDEFEISNEKYRKEKEDVLLSLYPNGELKKDOIIOPKBAFCITKILYSNOTSIIONCFPNAGSGVGPYLSQRPKLNNKCGECPILCTINKVNLKNIILBEEKLEKIKKEYEKERNESKVDNVLKRSKITEIKTYNGBEINKNSIIOILDNNNMKEDWENTYKDGKRNKFITHQSINMSPEFKTEYKNTPEISLNPDLKVVITYDNYVKNIDINTTGKRIILEIKNNKEXLTDITLPPQECQSCVLIHEKKIEIYGLNDEYKRAQBIYDKYISEEDAFGYIDVEYLOPEQEKRNKNKNNNGDENSGWEKNKEENIYMEGNKEENIYWEKNKEENIYWEGNKREBNIVAEGRKREBNIVAEENKNECNKINVASOMDNNKYIEDNSVOISBILKGCNRIEVDKRDNKLKLDNNKCVPFELQEKVDHNIFPSTSQNNSNLHTKEDIRDLNIELDNSPTVRESNDKEKIFSPVNYDCYINKSPIGENCKEKHIHSERFDLDTMYEGADTWGMLISMLWMTYI"
gene	complement(13072..13920) /gene="PF11545c" complement(13072..13920) /gene="PF11545c" /EC_number="3.4.25.1" /note="Similar to Nicotiana tabacum proteaseome delta chain precursor SW:PROD_TOBAC(P93395) (234 aa) fasta scores: E(): 3.3e-08, 28.689% id in 244 aa. Pfam match to entry PF00227 proteaseome, Proteaseome A-type and B-type, score 43.70, E-value 1.1e-10 and score 51.30, E-value 8.5e-13; ProfileScan hit to PS50247, Multispecific proteases of the proteaseome."
CDS	/codon_start=1 /product="proteaseome precursor, putative" /protein_id="CAD51995.1" /db_xref="GI:23505215" /db_xref="GOA:Q810U7" /db_xref="UniProt/TREMBL:Q810U7" /translation="MDVFNVSQIKCHEKSESDEVDIKTPISDGTIIIGIYDNGVMLACDSRTSSGTFISNKSCKRIKRNINENTYVCRSGASAHASOKLITIIIGHYVKNENRKKGRRHGEETIYDETYYDEBIDIDSINLDNNNNNDNLVTKNKFIEDRFVNDNPLEVNAVITIKLIITYNNPFCALIFGQYDKIKQOLYAVNNGSIIIEKEDFAVSGSGSITIOSYLOPKYKKFMWKCCFNULINCVMYAMHNDSGGILRIYNIYKSVFEEPTVNTQMNRQY"
gene	complement(join(15258..15314,15535..15628,15747..16056,16144..16285)) /gene="PF11550c" complement(join(15258..15314,15535..15628,15747..16056,16144..16285)) /gene="PF11550c" /codon_start=1 /product="hypothetical protein" /protein_id="CAD51996.1" /db_xref="GI:23505216" /db_xref="UniProt/TREMBL:Q812J7"

/translation="MGPELVKITSMLSHISEFFNYHISLPILIOYTLNKGSSNN
KIIVDEIISFLIKENIHNLDSKKKGYVEGYPNNLQAYSLKTLPHFVFLYADE
EYIKKYEEDNIAIFSYOKDYDINEBHNINIKPKQOVLISYRINISMLTIL
GNKKVNLHDFNDQMLIDHVKNEVSKNDENSTLNGDI"
join(17997, .18035,18047, .18112,18218, .18278,18587, .18845,
18040, .19172)
/gene="PF11555w"
join(17997, .18035,18047, .18112,18218, .18278,18587, .18845,
18040, .19172)
/gene="PF11555w"
/note="5 probable transmembrane helices predicted for
PF11555w by TMHMM2.0 at aa 28-45, 60-78, 85-107, 127-149
and 162-181"
/product="hypothetical protein"
/protein_id="CAD5197.1"
/db_xref="GI:23505217"
/db_xref="UniProt/TrEMBL:O812J6"
/translation="MTLKKIYNGHFRVRVKKKKKKKKKKYIYKVDGCSIHVS
LSEKADKSGFCRGYGVNIIYAGMFLYIIIMSCQKRFYIVHGFYLSIF
YLFPLHFRSMCMIPASYEKENVLSFCSFLDIFLCKHIIICVFIIFNSTEICKLN
VNIILHIYIYIYICATFVOKNK"
complement(join(19209, .19255,19418, .19474,19586, .22846,
23133, .23511))
/gene="PF11560c"
complement(join(19209, .19255,19418, .19474,19586, .22846,
23133, .23511))
/gene="PF11560c"

Query Match 9.6%; Score 68; DB 3; Length 281723;
Best Local Similarity 47.0%; Pred. No. 0.04;
Matches 247; Conservative 0; Mismatches 275; Indels 4; Gaps 1;

QY 1 AAGCTTTAAATTTGGGAAATTTAATATATATGATTAAGGCAATTAAGTTAA 60
DB 168592 AAAATTGAAAAATTCAAATTTAAAAAGGAAAAAATTTGA 168533
QY 61 TGAGAGTTTTTTTAAATTTTAAATTTTTCATGCAATGTTGTTTAA 120
DB 168532 TTGTATTTCTTTAATATATATATATATTTTTCATATTAATTTCCAT 168473
QY 121 TAAATTTTGACATTAACAAGTATTTTCCGCCAGATGAGTACGAGATGTC 180
DB 168472 TAATATATCTAATATGAAAAACAATAAATTTATTTTATATATATCAAGTA 168413
QY 181. CAGTCTTAAACCTGTAAGTTTGGTCTTACCAACCAATACCAAAAGTTAACG 240
DB 168412 ATATATATATTAATTAATGCAAGCATATTTCTTAATAGAAAAAAGGAA 168353
QY 241 ACCATACCGTTAATATATTTCTAACCGTTTAAAGTTTACATTAATCTACTAATC 300
DB 168352 AAAATATATATATTAATTAATA----TTTTTAAAGCAACATATATTTATATAAAC 168297
QY 301 CGCGTGTAATTAAGTTTATGAATGTGTTATTTGTAGTCACTGAAATTTAAT 360
DB 168296 CTTTAAATTAATATATCATATATATATATATATATATATATAATATATTAG 168237
QY 361 TTTTAACTGTTTCTTTTGGGTTCACTGTTACTTTTCTTTGACATCAA 420
DB 168236 TATATTTAATTAATTTATTTTATTTTCTCATATAGAAATATTTTTCATTGAAC 168177
QY 421 AATTAATGAGAGTGCCATATATGATGAGTGAATGAATGAATGAGTAATA 480
DB 168176 AATTATATGCAATATTTGATATATTAATTAAGGCTGAATATATATATTTATTGTAATA 168117
QY 481 AATTAATATGAAGTGAACAAAAAATTTAGTGAAGAAAGTTAACT 526
DB 168116 TATATATATATATATATATATATATATTAATTAATTAATTAATTAAT 168071

RESULT 9
AY527817/c 18556 bp DNA linear PRI 31-JAN-2004
LOCUS

DEFINITION Homo sapiens arachidonate 12-lipoxygenase (ALOX12) gene, complete cds.
ACCESSION AY527817
VERSION AY527817.1 GI:41324131
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 18556)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE Rieder, M.J., Daniels, R.L., da Ponte, S.H., Hastings, N.C.,
JOURNAL Ahearn, M.O., Rajkumar, N., Yi, Q. and Nickerson, D.A.
COMMENT Submitted (15-JAN-2004) Genome Sciences, University of Washington,
1705 NE Pacific, Seattle, WA 98195, USA
To cite this work please use: SeattleSNPs, NHLBI HL66682 Program
for Genomic Applications, UW-FHCRC, Seattle, WA (URL:
http://pga.gs.washington.edu).
FEATURES
source
1. 18556
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region
167, .237
/rpt_family="MIR"
variation
386 _type=dispersed
/frequency="0.56"
/replace="a"
864
/frequency="0.01"
/replace="c"
1138
/frequency="0.89"
/replace="a"
1158
/frequency="0.42"
/replace="c"
1440
/frequency="0.13"
/replace="a"
1586
/frequency="0.01"
/replace="a"
1649
/frequency="0.01"
/replace="a"
1661
/frequency="0.42"
/replace="c"
1764
/frequency="0.29"
/replace="a"
1947, .16598
/gene="ALOX12"
join(1947, .2114,2688, .2889,4372, .4453,4578, .4700,
4816, .4919,5169, .5329,6199, .6342,7465, .7674,11122, .11208,
11723, .11892,12351, .12472,15612, .15712,15821, .15991,
16109, .16598)
/gene="ALOX12"
/product="arachidonate 12-lipoxygenase"
join(1980, .2114,2688, .2889,4372, .4453,4578, .4700,
4816, .4919,5169, .5329,6199, .6342,7465, .7674,11122, .11208,
11723, .11892,12351, .12472,15612, .15712,15821, .15991,
16109, .16288)
/gene="ALOX12"
/codon_start=1
/product="arachidonate 12-lipoxygenase"
/protein_id="AAS00094.1"
/db_xref="GI:41324132"
/translation="MGRYRIIVATGAMLFSSGVNVOQLVGTREAEIEQLQRPARG
EEBEPDVAEDLGLQFRLKXKHLVDDAMFCDRITVQGGACAEVAFPCYRWVG
EDILSPEGTARLPDNLDMFQKREKELKDRQOIYCWATWKEGFLTIADRKODL

[illegible]

Center project name: zk178N6

----- Summary Statistics -----

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 223988 bases at least Q40

Consensus quality: 223337 bases at least Q30

Consensus quality: 223737 bases at least Q20

Insert size: 224728; sum-of-ctnigs

Insert size: 220016; 4.5% error; agarose-fp

Quality coverage: 9.09x in Q20 bases; sum-of-ctnigs Quality coverage: 9.29x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----

1 2652: contig of 2652 bp in length

* 2653 2752: gap of 100 bp

* 2753 111362: contig of 108610 bp in length

* 111363 111462: gap of 100 bp

* 111463 124950: contig of 13488 bp in length

* 124951 125050: gap of 100 bp

* 125051 128865: contig of 3815 bp in length

* 128866 128965: gap of 100 bp

* 128966 178608: contig of 49643 bp in length

* 178609 178708: gap of 100 bp

* 178709 217332: contig of 38624 bp in length

* 217333 217432: gap of 100 bp

* 217433 225328: contig of 7896 bp in length.

Location/Qualifiers

1. 225328

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone_id="DKEY-178N6"

/clone_id="Dkey-178N6"

1. 2652

/note="assembly_fragment:00004"

fragment_chain:1"

2753. 111362

/note="assembly_fragment:01763"

fragment_chain:1"

111463. 124950

/note="assembly_fragment:00219"

fragment_chain:1"

125051. 128865

/note="assembly_fragment:00045"

fragment_chain:2"

128966. 178608

/note="assembly_fragment:01061"

fragment_chain:2"

178709. 217332

/note="assembly_fragment:00426"

fragment_chain:2"

217433. 225328

/note="assembly_fragment:00093.0"

FEATURES

source

1. 225328

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone_id="DKEY-178N6"

/clone_id="Dkey-178N6"

1. 2652

/note="assembly_fragment:00004"

fragment_chain:1"

2753. 111362

/note="assembly_fragment:01763"

fragment_chain:1"

111463. 124950

/note="assembly_fragment:00219"

fragment_chain:1"

125051. 128865

/note="assembly_fragment:00045"

fragment_chain:2"

128966. 178608

/note="assembly_fragment:01061"

fragment_chain:2"

178709. 217332

/note="assembly_fragment:00426"

fragment_chain:2"

217433. 225328

/note="assembly_fragment:00093.0"

ORIGIN

Query Match 9 6%; Score 67.4; DB 2; Length 225328;

Best Local Similarity 36.3%; Pred. No. 0.058;

Matches 237; Conservative 0; Mismatches 412; Indels 3; Gaps 1;

Qy 6 TTTAAATTCGCAAAATTAATATATATATTAATGCAATAAAGTTAGATGACA 65

Db 111214 TATTTT 111273

Qy 66 GTTTTAAATTTTATATATATATATATATATATATATATATATATATATAT 125

Db 111274 TTTTAT 111333

Qy 126 TTTTGACATACCAAGATATATTTTCGCCAGCAATGAGCTACGAGAGATGCTGTC 185

Db 111334 TATATATATATTTTGTGTTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 111393

Qy 186 TTTAAACCTCGTAAAGTTTGGCTTACCCACCAATACCAACAAAGTAAAGACCAT 245

Db 111394 NNN 114453

Qy 246 ACCGTTAATATATCTAACCGTTTATAGTTTACATTAATCTACTATCCCGCT 305

Db 111454 NNN 11513

Qy 306 GTATTAAAGTTTATGAAATGCTGTTATTTGAGTCACGTAATTTATATTTT 365

Db 111514 ATATATATTTTATTTTATATATATATATATATATATATATATATATATATTT 111573

Qy 366 AGTACTTGT--TCTTTTGGGTCACTAGTACTTTTCTTGCATCAAAA 422

Db 111574 TTTATTTAATAAANAATATTTTNTTNTNNNAANNNTNNNNNNNNNNNNNNNN 11633

Qy 423 TTTATGACAGCTGTCATATATAGATGGAATGAATGATATGATATATAA 482

Db 111634 NNN 11693

Qy 483 TAAATATGAAAGTGAACAAATAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 542

Db 111694 TTTATTAATTAATTTATATATATATATATATATATATATATATATATATTT 111753

Qy 543 TTGGCTTGCAAGATTCATAGTTATTTTCAACCAAGAAAGAGTACCTTGCC 602

Db 111754 TTTTATTTTATTTTATATATATATATATATATATATATATATATATATATTT 111813

Qy 603 TCTCTCTCTTTTAAATGTTAACTGTTAAAGATGAGACTCAATTTCT 654

Db 111814 TATTT 111865

RESULT 12

CR847851 165797 bp DNA linear HTG 08-OCT-2004

LOCUS

DEFINITION

Danio rerio clone DKEY-115D7, *** SEQUENCING IN PROGRESS ***, 15

unordered pieces.

CR847851

ACCESSION

CR847851.3 GI:54019839

VERSION

HTG; HTGS_PHASE1.

KEYWORDS

Danio rerio (zebrafish)

SOURCE

Danio rerio

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 165797)

REFERENCE

Mclay, K.

Direct Submission

Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk

On Oct 9, 2004 this sequence version replaced GI:53828279.

----- Genome Center -----

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: zfish-help@sanger.ac.uk

----- Project Information -----

Center project name: zkp115D7

----- Summary Statistics -----

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 159489 bases at least Q40

Consensus quality: 160501 bases at least Q30

Consensus quality: 161321 bases at least Q20

Insert size: 164397; sum-of-ctnigs

Insert size: 191089; 11.8% error; agarose-fp

```
misc_feature      1. .16123
                  /note="assembly_fragment:01275"
                  fragment_chain:1"
misc_feature      16224. .26995
                  /note="assembly_fragment:00826"
                  fragment_chain:1"
misc_feature      27096. .29762
                  /note="assembly_fragment:00044"
                  fragment_chain:1"
misc_feature      29883. .32823
                  /note="assembly_fragment:00143"
                  fragment_chain:1"
misc_feature      32924. .59473
                  /note="assembly_fragment:01558"
                  fragment_chain:2"
misc_feature      59574. .77455
                  /note="assembly_fragment:01037"
                  fragment_chain:2"
misc_feature      77556. .80898
                  /note="assembly_fragment:00073"
                  fragment_chain:2"
misc_feature      81090. .84486
                  /note="assembly_fragment:00229"
                  fragment_chain:3"
misc_feature      84587. .87803
                  /note="assembly_fragment:00185"
```

ACCESSION	AC115599
VERSION	AC115599.2
	GI:28828893

KEYWORDS HTG.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
AUTHORS 1 (bases 1 to 125623)
Gloeckner, G., Eichinger, L., Szefanski, K., Pachbat, J., Dear, P.,
Lehmann, R., Baumgart, C., Parra, G., April, J. F., Guigo, R., Kump, K.,
Tunggal, B., Cox, E., Quail, M. A., Platzer, M., Rosenthal, A. and
Noegel, A. A.
Sequence and analysis of chromosome 2 of Dictyostelium discoideum
Nature 418. (6893), 79-85 (2002)
22093622
JOURNAL MEDLINE
PUBMED 12097910
REMARK The Dictyostelium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 125623)
AUTHORS Baumgart, C.
TITLE Direct Submision
JOURNAL Submitted (21-MAR-2002) Genome Analysis, Institute of Molecular
Biochemistry, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE 3 (bases 1 to 125623)
AUTHORS Baumgart, C.
TITLE Direct Submision
JOURNAL Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
Biochemistry, Beutenbergstr. 11, Jena 07745, Germany
4 (bases 1 to 125623)
Baumgart, C.
Direct Submision
Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
Biochemistry, Beutenbergstr. 11, Jena 07745, Germany
On Mar 4, 2003 this sequence version replaced gi:19570032.
CDS predictions from GeneID do not necessarily reflect true genes.
Further information is available from IMB Jena, Department of
Genome Analysis
(http://genome.imb-jena.de/dictyostelium/)
and the University Cologne, Institute for Biochemistry I
(http://www.uni-koeln.de/dictyostelium/project.shtml
Funding
Agency : Deutsche Forschungsgemeinschaft (DFG).
Location/Qualifiers
1. 125623
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="AX4"
/db_xref="taxon:44689"
/chromosome="2"
/map="4229038-4354721"
join<416..688,837..1349,1434..1493,1638..2931)
/note="Geneid exon scores (in order of location ranges) :
10.12, 21.67, 1.80, 118.27 - GSCJ_ID dd_00183"
/codon_start=2
/product="similar to Dictyostelium discoideum (S)ime
mold. GP138"
/protein_id="AA192328.2"
/db_xref="GI:28828944"
/translation="FKGNQKQOVNALLSQIYGVNYEGFQCLMNMVSGSDTVIGINI
FGVPTLVDSFVPELTYLFSSAISPSNIFSRLLPNTSPSEVFNIGKNDQIE
DIVLENTLVAFKISVPLPAFSTGATGILLINNAGPALPQOPONNPKLOL
LNIINGFRPSNAGSLSNLTETLNTNNYTKRTTFPPFPNQTAKAINPHSADP
LNQLFEPFPINNINPLELTLNENLGNISNDVLTNLSKNLKLTESNHLLPNKRY
PLNCSLDLSNYSVAGLSDVFNRLTSLSNVYGENLPASNPDEPKLTYLLLSNQF
NLSLPEEYCSFVKSGIADGSLDTSVPCFCGCEPGLFPNNFDDFPDTPADV
CAFSLDVFNQDLKTNETTVEFTGNSNGMSDIESIQANVELAIEBNKKIKITIP
PGAGIOMATLKFGNSPSIEVIINERYEPTIDSYVVGSGIFPFGSYFSYNDYENIF
TINKYKTAHQTIEDONLNGVCGENDTPISVUKREAFNVSIDVAGKQSEVTFE
ITFGNTMLSDLSISQNLNTGDFSGFGCTTDFSNPIKINDIVIDQILLETETR
INITYPPIENAGTYSLYIEVGDFSEFNEIKFTETPPPTPTPTPTPTSTSTST
SDDEGLSTPTLSISRTYLSIILLINFYF"
complement(1013950..4537,4680..4808)
/note="Geneid exon scores (in order of location ranges) :
38.84, 6.50 - GSCJ_ID dd_00185"
/codon_start=1
/product="similar to Ralstonia solanacearum (Pseudomonas
solanacearum). Putative GST-related protein (EC 2.5.1.18)"

/protein_id="AA051485.1"
/db_xref="GI:28828894"
/translation="MQNKFDEFTIINDHKSPTYKVELMLIELGVSYESISINDYLN
NSNNRIPALIDMTFPKVPNTLYESGVALYLAKFSLLDLPFLNATLONNSQITIS
VMOVLNLTSPNN
ADISVWLLHLLKNSNSQITISRNINKMELINORDSIKLANNNNNNNNNNN
NANNNGIKFSNDGLAINL"
complement(5354..7216)
/note="Geneid exon scores (in order of location ranges) :
110.35 - GSCJ_ID dd_00186"
/codon_start=1
/product="similar to Plasmodium falciparum (isolate 3D7).
Hypothetical protein"
/protein_id="AA051486.1"
/db_xref="GI:28828895"
/translation="MGFTMEYENNRKQDGVDDNNNSINNKTPSNNNNNN
YSSSSNNNNNNNDYSDNDYNOITPIQONDYQDSTNNNNNSNNNDYNOIPYLN
NDFQDTTNNNSYTYDYNNTYNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NOINNSNN
IYKNSNLGNKNIDNEKSNKYKPLKXDSYTIHKOIANNI VLYFKDAGYQSDL
IAIKDSKTDILKEIKYFVGLDYKESISTLLFKMI INDKTAKPHEFYNSPSETI
LPIESDPELLKPNHLIIOGSPFIHRNRKMPPTFTSTYPTGTTEYFKYQKDPV
KIDRKACPYVIDFKKSDSIPOYGFYKGVPSQITNYSPIRPEKCPYIYHND
NNSERNATWLKNTFLIFRYLQSRANDGTATYGOQLPNLVSQSSISILBPLKDSI
NOSKLCIGRSTLSFGNPLCNTHYSKDLMDAEKCKPYFLDANNEPIQLPKDQES
FKENKKEKKTWFIKDFAYIDPSNLOSITDKVKVRYTQENNNYNSYDYDY"
complement(7921..9687)
/note="Geneid exon scores (in order of location ranges) :
222.87 - GSCJ_ID dd_00188"
/codon_start=1
/product="similar to Agrobacterium tumefaciens (strain C58
/ ATCC 33970). 3-methylcrotonoyl-CoA carboxylase beta
subunit"
/protein_id="AA192331.2"
/db_xref="GI:28828896"
/translation="MLKSIISLKNQIILKNIIINGRIINNVGKLSKSLKINYS
STTDRTFILDGTIDKNSAIEYKONLINNSTIKQIKENIKLIGGGEKLANQINSRG
KLIVBERIALIDVGSPLFESQALMGYGEVAAGIITIGIGIHEVCEYVND
STVKGGTPIPIVKKHILROEIOAENNLPCITLVGSGANLPRDVPDRHFGFIF
PNOAMSAKRIQIIVWMSCTAGAYYPAMADESVIYKGTITPILGCGPVLKATGE
IYTSBELGADJHCTSGVTDHYARDDBAIALITRIYSNLRKQPSVITETEPPL
YPTSLAGIYPSDLKKNPDIRVIALRVIAVLDGDFEKLXGTLLCGFARVHGMPEI
IANNGIILPSESIVKGAHFIELCNQIGIPLVFNQNTGFWGVTYESGIAXKAGAKVM
AVATAKVPKITWIIIGSGFAGNYGMCGRSFRPLTYPMPNNAKISVNGGQAASVLAOI
OKDNMAKENKOWSPREENTFKKPIIDKPEEBSIYSSARCMDDVIDPDQRKIAL
SLSACMGPINPSPSGFVFRK"
complement(9983..11383)
/note="Geneid exon scores (in order of location ranges) :
98.71 - GSCJ_ID dd_00191"
/codon_start=1
/product="similar to conserved hypothetical protein
[Schizosaccharomyces pombe]"
/protein_id="AA192333.2"
/db_xref="GI:28828897"
/translation="KRATHSREVDVKTETINIKETSTTKVTESAIDRFKSKSKTK
KHKHVKVNSDNDNDNNINNERKVKKSTLSLKNKKPSIKKDLBPVNDORDT
AKRAEKLTYENRPLBEGNPEYNNKLETOYKRYKUNGLERLRLSLVDINKLSRYR
CYRDTTILKRLBEYTKETSILNKGELVEITFGSSSQPLAKSKSDVDLWVSFEFLT
KNDRTTKTCYQSSILIRANGFYNIKPIIHAKPIYKFPDPKTEPCDDTLTDSGNTG
VYKECOLPILPIVLIIFCKMASVLANINDNSOGTSSYISNNVI FVLOKGLPSY
KOIGFKNSSKEVQVYVNNRDADLLVTFKFGYIFDRNMMNIIADNNDKKEPM
AEEELPNNREYDKNNHYLPVMDKDLDDTDLFFIKDICIAGHNTTRCIRLSQPLTLI
KYFQKAYANLISGKLP"
13531..11383
/note="Geneid exon scores (in order of location ranges) :
16.95 - GSCJ_ID dd_00194"
/codon_start=1
/product="hypothetical protein"
/protein_id="AA192334.2"
/db_xref="GI:28828898"
/translation="WSIITYSPVPQLFEVELRDLDVNSEEFKFAQIFNNGVPSIK
KIOVVYNSVWKNYKNGNLGNSLKEIFVHSGRLNPKLIYTNGLIVERTSAGLYFA"
14573..14956

```

/note="gened exon scores (in order of location ranges) :
5_39 - GSC1 id dd_00176"
/codon_start=1
/produc="hypothetical protein"
/product_id="AA029335.2"
/db_xref="GI:2882889"
/translation="MNSSEFVKKIQVYNNKKYWDNYIQHMLKSNHSLSEELFVFGTSLN
DPSLIYNGLAERKQSGLYFAIKSESGSFYKNFDQSQIFCILIPRONVSPRFV
IKNDNHAPQYALISNSKRYNSIML"
join(15573..16125,16226..17482)
/note="gened exon scores (in order of location ranges) :

```

```
CDS      18511..20217
          /note="GeneID exon scores (in order of location ranges):
          169.96 - GSCJ_ID dd_00184"
```

```

/ctansaction="MSVILINSIADLKTEPESTVTTVVVKAFTVNINIGGDFYYDSD
DTTATVDNATVTVNMGKKAKKATVSDPNDANVTHTGALKDGVTCOSACITMHWXSVK
VAPKGIKRPAPGTFKISGIDILNNTSIAPHRMGTSTVAAGIFPATLLISDQGTGMWVK
KARWEISNLIYYGENDKQVNTKGFQOMTELEGEFIRTNINFRNGGVCAVSMWDTLD
TLDQVYVTRCTGAGVYGTWGSATSGSMHDLAEVJTNFINONCTGAINPLOCRTOSD
IIRBNGTIHSDDPGLSNQWVTEALSMEDCCNPLNTRCYVITIKOKLOSTSLNTRN
DSTKSTWLSIWENGQTEIENVGIRRTGSGHNGEALTEVOKSPQVONSGWGRGSCFPL
SODITNLIIVGSLGFSLLTNSHDVAOSGHHNGEALTEVOKSPQVONSGWGRGSCFPL
TDMVYLPANKNKSIIYIKLGPYNNVTVALIETSDARBYAGVCFRMWYDQTKLYKTEAT
VTTIETIVALPQASNMACKNGLAIGSDVEPITVPIGTNDQPIYINGSLYKATAVV
KQ".
21626
/notes="Geneid exon scores (in order of location ranges) :
6..32 - GCGC_ID dd_00190"

```

Query Match	9.5%;	Score 67;	DB 3;	Length 125623;
Best Local Similarity	48.0%;	Pred. No. 0.078;		
Matches 254;	Conservative	0;	Mismatches 270;	Indels 5; Gaps 2;
QY	9	AAATTTGGGAAAATTTTAAATATATGATTAATTAAGGCAATTAAGTTAGATGAGATT	68	
Db	26945	AACATACAGATCTATATTTTCATATATTTTAAAGAAATTAATCAATGTGGAATTAGAGA	27004	
QY	69	TTTTTAAATTTTATTAATAAAATTAATTTTTCATGAAATGTTTAAATTAATTT	128	
Db	27005	ATTGTGTAATTAATTTAAACCAATGATTTTATTAATTAATCAATAAGATTAATTTGATTTGTT	27064	
QY	129	TGACATACCAAGTATTATTTTCC--GCCAGAAATGAGCTACGAGACATGCTCGTCT	186	
Db	27065	TCATCAAAATTAAGTAATATTTTCCATATTTATCAAAATTAATTAATCTAAAGTGTAAATCA	27124	
QY	187	TTAAACCTCGTAAAGTTTGGTCTTAACCAACCAATPACCAAGTAAGTAAGACATTA	246	
Db	27125	TTTGAAATGTGGAATTTTACAAAAAAATTCACATTAAGAAATCTTTATCAAAATCA	27188	
QY	247	CCGGTTAATATATCTTAACCGGTTTAAATTGATTAATCAATTAATCAATCCCGCTG	306	
Db	27185	GTTAATTTTAAAAAAGAAACGAGTTCAAAATTCAAATCTTGATATATGATTTTCAAAATGGT	27244	

[illegible]

RESULT 14	CR753884/c	LOCUS	DEFINITION	CR753884	148394 bp	DNA	linear	HTG 04-SEP-2004
ACCESSION	CR753884		Danio rerio clone CH211-276D21, WORKING DRAFT SEQUENCE, 8 unordered pieces.	CR753884				
VERSION	CR753884.4		GI:51890093					
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT.							
SOURCE	Danio rerio (zebrafish)							
ORGANISM	Danio rerio							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.							
REFERENCE	1 (bases 1 to 148394)							
AUTHORS	McLay, K.							
TITLE	Direct Submission							
JOURNAL	Submitted (03-SEP-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK. E-mail enquiries: genbank@sanger.ac.uk							

```

COMMENT
On Sep 4, 2004 this sequence version replaced gi:51870358.
-----
Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
-----
Project Information
Center project name: zc276bd21
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: dye-terminator; 100% of reads
Consensus quality: 143152 bases at least Q40
Consensus quality: 144400 bases at least Q30
Consensus quality: 145119 bases at least Q20
Insert size: 147694; sum-of-contigs
Insert size: 154861; 2.1% error; agarose-fp
Quality coverage: 10.33x in Q20 bases; sum-of-contigs Quality
coverage: 10.81x in Q20 bases; agarose-fp
-----
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 2608: contig of 2608 bp in length
*
* 1
* 2609
* 2708: gap of 100 bp
*
* 2709
* 28743: contig of 26035 bp in length
*
* 28744
* 28843: gap of 100 bp
*
* 28844
* 32353: contig of 3510 bp in length
*
* 32354
* 32453: gap of 100 bp
*
* 32454
* 44987: contig of 12534 bp in length
*
* 44988
* 45087: gap of 100 bp
*
* 45088
* 47132: contig of 2045 bp in length

```


*	109983	gap of 100 bp
*	109984	contig of 8173 bp in length
*	118157	118156: gap of 100 bp
*	118257	118256: gap of 100 bp in length
*	129072	contig of 10816 bp in length
*	129173	gap of 100 bp
*	182677	contig of 53505 bp in length
*	182678	gap of 100 bp
*	182679	contig of 13985 bp in length
*	182778	182777: gap of 100 bp
*	182779	contig of 13985 bp in length

FEATURES

```

misc_feature
1. .5760
/note="assembly_fragment:00009
fragment_chain:1"
5861. .20214
/note="assembly_fragment:00500
fragment_chain:1"
20315. .33680
/note="assembly_fragment:00347
fragment_chain:1"
34781. .43318
/note="assembly_fragment:00123
fragment_chain:1"
43419. .73825
/note="assembly_fragment:00837
fragment_chain:1"
73826. .109883
/note="assembly_fragment:01226
fragment_chain:1"
109984. .118156
/note="assembly_fragment:00051
fragment_chain:1"
118257. .129072
/note="assembly_fragment:00216
fragment_chain:2"
129173. .182677
/note="assembly_fragment:01642
fragment_chain:2"
182778. .196762
/note="assembly_fragment:00664.0"

```

ORIGIN

Query Match	9.4%;	Score 66.6;	DB 2;	Length 196762;	
Best Local Similarity	48.5%;	Pred. No. 0.08;			
Matches 242;	Conservative	0;	Mismatches 254;	Indels 3;	Gaps 2.

Qy	11	ATTGGGAAATTTAATAATATACGTATTAATGGCAATTAAGAATGAGACTT	70
Db	118101	ATTTPAAAATAATTATTAATAATTATTAATTAATAAAAAATMAATTTATATTTAT	118042
Qy	71	TTTAATTTTTATATATAAATAATTTTTTCATCATGAATTGTTTTNAGATAAAATTTG	130
Db	118041	TTTTTATTTTATATAATTAATTAATTTAATATATATATTTTTTATTTATATATAT	117982
Qy	131	ACATTAACAAGATATATTTTTCCGCCAGCAATTGAGCTACGAGAGTGCCTCTTTAA	190
Db	117981	ATTTAATATTTTATATATATATTTATTAABAATPATTAATATATATATATTAATA	117922
Qy	191	ACCTCGTAAGTTTGGTCTTACCACCAACCAATACCCACAAAGTAAAGACCAATACCG	250
Db	117921	TTTTAT - AATTAATTAATTAATTAATTTAATATTAATAAATTTATATATATATTA	117866
Qy	251	TTAATTAATTTCTAACCGTTTATTAAGTTTACATAATCATTACTAATCCGGGTGAAT	310
Db	117863	TTAATATTAATTTTATTAATTAATTAATTAATTAATAATTAATTAATTTATATTTAT	117804
Qy	311	TAACTTTATGAAACTGGTATTTTGTAGGTACGCGAAATTATTAATTTTGTATAC	370
Db	117803	TAAATTAATATAATAATATATATTTTAAATTAATATTTATTTAATTTAATATTTAA	117744

QY	371	TTGTTTTTCTTTTTTGGGTTCAACIAGTACTTTTTTCCTTGACATCAAAATTAAATGTA	430
Db	117743	TAATAATTAATTAATTAATAAAATTTTAAATAAATTAATTAATTAATAAATAAATTAATTTAT	117684
QY	431	GACCAAGTGTCCATATATAGATGTGGAATGAATGAATATATGAGTAATTAATTAATATA	490
Db	117683	TAATATATATATAAATATATA-TAATATATATATATATATATATATATATATATATATATA	117625
QY	491	GAAGTGAACAAAAAATAAT	509
Db	117624	TATATATATATATATATAT	117606

Search completed: November 17, 2005, 16:54:15
Job time : 3275 secs

QY		491	GAAAGTGAACAAAAAAT	509
Db	117624	TATATATATATATATAT	117606	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2005, 13:43:52 ; Search time 473 Seconds
(without alignments)
8823.293 Million cell updates/sec

Title: US-10-810-788a-4

Perfect score: 705
Sequence: 1 aagcttcaattcgggaaa.....tctgtggtcacttgatacc 705

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseq1980s:*
2: geneseq1990s:*
3: geneseq2000s:*
4: geneseq2001as:*
5: geneseq2001bs:*
6: geneseq2002as:*
7: geneseq2002bs:*
8: geneseq2003as:*
9: geneseq2003bs:*
10: geneseq2003cs:*
11: geneseq2003ds:*
12: geneseq2004as:*
13: geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64.2	9.1	8056	ABZ10246	ABZ10246 Haematopo
2	63.4	9.0	612	AAH71471	Aah71471 Human cer
3	62.6	8.9	8056	ABZ10246	ABZ10246 Haematopo
4	62.4	8.9	531	ACNS1594	ACNS1594 Cotton an
5	62.2	8.8	8056	ABZ10100	ABZ10100 Haematopo
6	60.6	8.6	11422	ABK31495	ABK31495 Signal tr
7	60.6	8.6	11422	ABK31495	ABK31495 Signal tr
8	60.4	8.6	560	ABL32218	ABL32218 Human imm
9	60.2	8.5	626	ACN47472	ACN47472 Cotton pr
10	60.0	8.5	499	ABV60941	ABV60941 Human pro
11	59.4	8.4	8056	ABZ10100	ABZ10100 Haematopo
12	59.4	8.4	15954	ABK31495	ABK31495 Signal tr
13	59.4	8.4	15954	ABK31495	ABK31495 Signal tr
14	59.2	8.4	6816	ADQ24856	ADQ24856 Human soc
15	59.2	8.4	50000	ABL55644	ABL55644 Human soc
16	59.0	8.4	469	ACN62049	ACN62049 Cotton gy
17	58.8	8.3	752	ADL44766	ADL44766 Human ova
18	58.6	8.3	428	ACNS2919	ACNS2919 Cotton an
19	58.6	8.3	469	ACN62049	ACN62049 Cotton gy
20	58.6	8.3	547	ACN62109	ACN62109 Cotton gy

C	21	58.4	8.3	631	5	AD173251	AD173251 Human ova
C	22	58.4	8.3	631	5	ADL38382	ADL38382 Human ova
C	23	58.2	8.3	554	13	ACNS1258	ACNS1258 Cotton an
C	24	58.0	8.2	9646	6	ABL33688	ABL33688 Human imm
C	25	57.8	8.2	494	5	ABV10021	ABV10021 Human imm
C	26	57.8	8.2	7823	4	AA545490	AA545490 Chemical
C	27	57.8	8.2	7823	4	ABL34061	ABL34061 Human imm
C	28	57.8	8.2	7823	6	ABK31493	ABK31493 Signal tr
C	29	57.8	8.2	7823	6	ABK31493	ABK31493 Signal tr
C	30	57.6	8.2	353	13	ACNS3258	ACNS3258 Cotton an
C	31	57.6	8.2	486	13	ACN61227	ACN61227 Cotton gy
C	32	57.4	8.1	550	5	ABV57072	ABV57072 Human pro
C	33	57.4	8.1	365	13	ACNS6344	ACNS6344 Cotton an
C	34	57.4	8.1	4857	13	AD589407	AD589407 Oligonuc
C	35	57.4	8.1	7857	6	ABO67075	ABO67075 Human ang
C	36	57.4	8.1	9001	13	AD589773	AD589773 Oligonuc
C	37	57.2	8.1	353	13	ACNS3258	ACNS3258 Cotton an
C	38	56.8	8.1	8423	6	ABL33407	ABL33407 Human imm
C	39	56.6	8.0	376	5	ADL44697	ADL44697 Human ova
C	40	56.6	8.0	574	13	ACNS6366	ACNS6366 Cotton an
C	41	56.4	8.0	385	5	ADL37765	ADL37765 Human ova
C	42	56.4	8.0	385	5	ADL72626	ADL72626 Human ova
C	43	56.2	8.0	17848	4	AA545323	AA545323 Chemical
C	44	56.2	8.0	17848	4	ABK39976	ABK39976 Human che
C	45	56.2	8.0	17848	6	ABK28164	ABK28164 DNA trans

ALIGNMENTS

RESULT 1	ABZ10246/c	ABZ10246 standard; DNA; 8056 BP.
ID	ABZ10246	
XX	ABZ10246;	
AC	ABZ10246;	
XX	16-JAN-2003 (first entry)	
XX	Haematopoietic cell proliferation disorder related DNA sequence #386.	
DE	Human, hematopoietic cell proliferation disorder; cytostatic;	
KW	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;	
KM	cytosine methylation state; gene; ds.	
XX		
OS	Homo sapiens.	
XX		
PN	W0200277272-A2.	
XX	03-OCT-2002.	
PD		
XX	26-MAR-2002; 2002WC-EP03401.	
PF	26-MAR-2001; 2001US-0278333P.	
XX		
PR	26-MAR-2001; 2001US-0278333P.	
XX		
PA	(EPIC-) EPICENOMICS AG.	
XX		
PI	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;	
PI	Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu F;	
PI	Levin A, Lipscher E, Mader S, Model F, Mueller V, Otto T, Pelet C;	
PI	Schwope I, Ziebarth H;	
XX		
DR	WPI; 2003-018942/01.	
XX		
PT	Detecting and differentiating between hematopoietic cell proliferative	
PT	disorders, comprises contacting a target nucleic acid with a reagent that	
PT	disturbes, comprises contacting a target nucleic acid with a reagent that	
XX	disturbes, comprises contacting a target nucleic acid with a reagent that	
XX	disturbes, comprises contacting a target nucleic acid with a reagent that	
PS	Claim 28; SEQ ID NO 366; 117bp; English.	
XX		
CC	The present invention describes a method for detecting and	
CC	differentiating between hematopoietic cell proliferative disorders	
CC	associated with at least 1 gene and/or their regulatory regions in a	
CC	subject. The method comprises contacting a target nucleic acid in a	

biological sample obtained from the subject with at least 1 reagent, which distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. ABZ09861 to ABZ1118 CC represent specifically claimed nucleotide sequences from the present invention. Oligonucleotides from the present invention can be used: for differentiating between healthy haematopoietic cells and proliferative disorder haematopoietic cells; for differentiating between acute lymphocytic leukaemia and acute myelogenous leukaemia; as probes for determining the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) of haematopoietic cell proliferation disorder related sequences and their complements; and as primers for the amplification of haematopoietic cell proliferation disorder related DNA sequences. The nucleotide sequences from the present invention can also be used for detecting a predisposition to, differentiation between subclasses, diagnosis, prognosis, treatment and/or monitoring of haematopoietic cell proliferative disorders. The present method enables a highly specific classification of haematopoietic cell proliferative disorders allowing for improved and informed treatment of patients

Query Match	9.1%	Score	64.2	DB	8	length	8056
Best Local Similarity	45.7%	Pred. No.	0.022				
Matches	260	Conservative	0	Mismatches	308	Indels	1
						Gaps	1

QY	5	TTTTAAATTTGGGAAAAATTTATATATAATGTAATTAATGGCAATTAAGATTAAGTACATGAG	64
Db	7154	TTACAAATTATATATATATATATATTTTATTAATATTTAAAAAAATTTATAAATTTATTAAT	7095
QY	65	AGTTTTTTTAAATTTTATATATAAAATATTTTTTGCATGAATGTGTTTTAAGATAAA	124
Db	7094	TAAAAAATTAATATTTAAAAAATATATAATTTTAATTAATTAATAATTTTTTTTTTTT	7033
QY	125	ATTTTGACATACCAAGTATATTTTCCGCCACGAATTGAGCTACGAGAGTGTCTGT	184
Db	7034	AAAAAAAATCAAAAATTAATATTTTTTTTTTTTTTTTTTAAATTAATAATCAATATATTTTT	6975
QY	185	CTTTAAACCTCGTAAAGTTTGGCTTACCCAAACCATTAACCAAGTAAGTAAACGACA	244
Db	6974	TATATATATATATATTTTTTAAATTAATAAAAAATTTTAAAAAATTAATAACA	6915
QY	245	TACCGGTTAATATATTTCTAAACGGTTATATAGTTTACATTAATCATTTACTAATCCCG	304
Db	6914	TTCAATTTATTCAAAAAATTAACAATTTTTTTTAAAAAATTTTATTAATTT	6855
QY	305	TGTAATTAAGTTTATGAAATGTGGTATTTTGTAGTCAAGTGAATTTATTAATTTT	364
Db	6854	TTTCAATTCATTTATTAATAATTTTTTATTTATTTTAAAAAAATTCATTTTCACAATTC	6795
QY	365	TAGTACTGTCTTTCTTTTTGGGTCACAAGTACTTTTTCTTGACATCAAAAT	424
Db	6794	AAATATTTTTTATTTTTTTTATATTTTTTTCATTTAATATATTTTTTAAAAATCA	6735
QY	425	ATTGTAGACGAGTGCATATATAGATGTGAATGAATGAATATTTAGATTAATAATA	484
Db	6734	ATATTTCAAATTTATATATTAATCAACATTAATAAACATAAAAAATTTATTTTTTTAA	6675
QY	485	AATATAGAAATGAACAAAAAAATTAG-TGGAAAAGGTAACGTGAAAAGAAAAGGCTTAT	543
Db	6674	AATTTAATTAATAAAATTAATAATTTCAATATAAAAAACATAAAAATTAACAATAAATAT	6615
QY	544	TGGCTTGGCAAGATTCCTAAGTTATTT	572
Db	6614	AAATTAATTAATAAAATTAATAATTTTTTTT	6586

RESULT 2	
AAH71471/C	standard; cDNA; 612 BP
ID	AAH71471
XX	
AC	AAH71471;
XX	
DT	19-SEP-2001 (first entry)

XX	Human cervical cancer marker nucleic acid 2745.
DE	
XX	
XX	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200142467-A2.
XX	
PD	14-JUN-2001.
XX	
PF	08-DEC-2000; 2000WO-US033312.
XX	
PR	08-DEC-1999; 99US-0169681P.
PR	21-DEC-1999; 99US-0171350P.
PR	14-MAR-2000; 2000US-0189315P.
PR	12-MAY-2000; 2000US-0203791P.
PR	09-JUN-2000; 2000US-0210600P.
PR	21-JUL-2000; 2000US-0220114P.

PT New isolated nucleic acid for diagnosing and treating cervical cancer and
PT for assessing and detecting compounds for treating the cancer.

PS Claim 1; Page 564; 1051pp; English.

The invention relates to novel genes (AAH68727-AAH73383) associated with cervical cancer with cytostatic activity. The nucleic acids and encoded polypeptides are useful: to assess if a patient is afflicted with cervical cancer or has a pre-malignant condition; to monitor the progression of cervical cancer or a premalignant condition in a patient; and to select and/or assess the efficacy of a compound or therapy for inhibiting cervical cancer in a patient. The nucleic acids may also be useful for gene therapy.

SQ Sequence 612 BP; 380 A; 50 C; 2 G; 178 T; 0 U; 2 Other;

Query Match Similarity	9.0%	Score 63.4	DB 4	Length 612
Best Local Similarity	52.7%	Pred. No. 0.029		
Matches 136	Conservative 0	Mismatches 122	Indels 0	Gaps 0

Qy	251	TTAATAATATCTCAACCGGTTATAGTTACATAAATCATTTACTAATCCGCGGTAAAT	311
Db	265	TTTTTTTTTTTAAAATAATTTTTTGGTATGTTAAATAATGGGATGGGTATTTT	206
Qy	311	TAAGTTTATGAAATGCTGTTATTTTGTAGGTCACGTAAATTTATATTTTGTAC	370
Db	205	TTATTTTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAT	146
Qy	371	TTGTTTTTCTTTTTTGGGTCACAGTACTTTTTTCCTGACATCAAAATATATGTA	433
Db	145	TTTAAATAAAAAATTTAAA	86
Qy	431	GACAGATGGTCATATATAGATGTGAATGAATGAAATATGTATTAATTAATATA	490
Db	85	AAGAAAAA	26
Qy	491	GAAGTGCAACAAAAAA 508	
Db	25	AAAAAAAAAAAAAAAAAAAA 8	

RESULT 3	
ABZ10246	
ID	ABZ10246 standard; DNA; 8056 BP
XX	
AC	ABZ10246;
XX	

DT	16-JAN-2003	(first entry)
XX		
DE	Haematopoietic cell proliferation disorder related DNA sequence #386.	
XX		
KW	Human; haematopoietic cell proliferation disorder; cytostatic;	
KW	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;	
KW	cytosine methylation state; gene; ds.	
XX		
OS	Homo sapiens.	
XX		
XX	MO200277272-A2.	
PX		
PD	03-OCT-2002.	
XX		
PF	26-MAR-2002; 2002WO-EP003401.	
XX		
PR	26-MAR-2001; 2001US-027833P.	
PA	(EPIG-) EPIGENOMICS AG.	
PX		
P1	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J,	
P1	Olek A, Pieglenbrock C, Adorjan P, Grabs G, Leische R, Lau E,	
P1	Lewin A, Lipscher E, Mater S, Model F, Mueller V, Otto T, Pelet C,	
P1	Schwope I, Ziebarth H;	
XX		
DR	WPI; 2003-018942/01.	
XX		
PT	Detecting and differentiating between hematopoietic cell proliferative	
PT	disorders, comprises contacting a target nucleic acid with a reagent that	
PT	distinguishes between methylated and non-methylated CpG dinucleotides.	
PS	Claim 28; SEQ ID NO 386; 117pp; English.	
CC		
CC	The present invention describes a method for detecting and	
CC	differentiating between haematopoietic cell proliferative disorders	
CC	associated with at least 1 gene and/or their regulatory regions in a	
CC	subject. The method comprises contacting a target nucleic acid in a	
CC	biological sample obtained from the subject with at least 1 reagent,	
CC	which distinguishes between methylated and non-methylated CpG	
CC	dinucleotides within the target nucleic acid. AB209861 to AB21118	
CC	represent specifically claimed nucleotide sequences can be used; for	
CC	invention. Oligonucleotides from the present invention can be used; for	
CC	differentiating between healthy haematopoietic cells and proliferative	
CC	disorder haematopoietic cells; for differentiating between acute	
CC	lymphocytic leukaemia and acute myelogenous leukaemia; as probes for	
CC	determining the cytosine methylation state and/or single nucleotide	
CC	polymorphisms (SNPs) of haematopoietic cell proliferation disorder	
CC	related sequences and their complements; and as primers for the	
CC	amplification of haematopoietic cell proliferation disorder related DNA	
CC	sequences. The nucleotide sequences from the present invention can also	
CC	be used for detecting a predisposition to, differentiation between	
CC	subclasses, diagnosis, prognosis, treatment and/or monitoring of	
CC	haematopoietic cell proliferative disorders. The present method enables a	
CC	highly specific classification of haematopoietic cell proliferative	
CC	disorders allowing for improved and informed treatment of patients	
XX		
SO	Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;	
	Query March	8.9%; Score 62.6; DB 8; Length 8056;
	Matches Local Similarity	45.0%; Pred. No. 0.043; Indels 1; Gaps 1;
	Matches 274; Conservative 0; Mismatches 334;	
OY	11 ATTTCGGAAATTTAATATATATATATATGAATGCAATATAAGTATGATGAGATT	70
DB	2109 ATTTTAATTAATTTTAAATTTTTTTTATTTATTTAAATAAAAAATAAAAAAA	2166
OY	71 TTDAATTTTATTAATAAAAAATATTTTTCATGCAAAATGTGTTTAAAGATMAAATTTG	130
DB	2169 ATTATTAATTTTATTAATAAAAAATTAATAAAAAATTAATAAAAAATTAATTT	2228
OY	131 ACATACCAATATATATTTTCCGCACGAATTGAGTCTACAGAGAGATGTCCTGCTTAA	190
DB	2229 AATTTTAATTAATAAAAAATTTATTTTTTTTAAATTTTAAATTAATTAATTTTAAATTTT	2288

Oy	191	ACCCGGAAGGTTTGGCTTACCCCAACCCCAATCCCAAGGTAACGACCATTCGCG	250
Db	2289	AATTTATTTATTTTAAATAATTTATTAATTTTGTATTTTAAATTTTATTTAATT	2348
Oy	251	TTAAATAATTCCTAACCGGTTTATTAAGTTTACATTAATCATTAATCCGCGTAAAT	310
Db	2349	TTTAAATATATTTTATTTATTTTATTAATAATTTATTAATTTTATTTTGAATAAT	2408
Oy	311	TAACTTTTATGAATATGCGTATTTTGTAGTCACGGAATTATTAATTTTATAGTAC	370
Db	2409	TAAAAAATAATTTATTTATTTTAAAAATTTAAATTTTATTTTATTTATTTA-TAT	2467
Oy	371	TTGTTTTCTTTTTTGGGTCACACTAGTTACTTTTTCTTTGCATCAAAATTAATGTA	430
Db	2468	ATTTTATTTATTTTGTGTTTAAATAATTTAAATAATTTAAAGAAAATTTAAAAATTA	2527
Oy	431	GACGAGTGTCATATATTAAGTGTGAATGAATGAATATTGATATTAATTAATATA	490
Db	2528	TTAAATATTAATAATTTAAATTTAAAAATTTTAAATTTTAAATTTTAAAAATTA	2587
Oy	491	GAAAGTGAACAAAAAATAATAGTGAAGAAAGTAACTGGAAGAATAAGCTATTTGGCTG	550
Db	2588	AATTAATGAATAAAAAAATTAATTAATTAATTAATTTAAAAAATAAAAAAATA	2647
Oy	551	GCAAGATTTCATTAAGTTTATTTTCACAAAAAGAAAGATTAAGTGGCTTGTCTCTC	610
Db	2648	GTTTAAATTTTAAATTTTAAAAAATAATTAATTTTAAATTTTAAATTTTATTTATTTT	2707
Oy	611	TCCTTTTAAA 619	
Db	2708	TTTTTTTAA 2716	
RESULT 4			
ACNS1594			
ID	ACNS1594	standard; cDNA; 531 BP.	
AC	ACNS1594;		
XX	02-DEC-2004	(first entry)	
DT			
XX			
DE			
XX			
KM		Cotton androecium tissue EST Clone ID: LIB3828-004-Q1-N6-H2, SEQ:6375.	
KM		Cocton; plant; EST; expressed sequence tag; transgenic plant; androecium;	
KM		variety Nuccio033b; library LIB3828; molecular tag; molecular marker;	
KM		genetic mapping; molecular mapping; seed germination; plant growth;	
XX		plant quality; plant yield; plant breeding; tissue printing; ss.	
OS		Gossypium hirsutum.	
XX			
PN		US2004123340-A1.	
XX			
PD		24-JUN-2004.	
XX			
PF		12-DEC-2001; 2001US-00021323.	
PR		14-DEC-2000; 2000US-0255619P.	
XX			
PA		(DEIK/) DEIKMAN J.	
PA		(FENG/) FENG P C C.	
PA		(FINC/) FINCHER K L.	
PA		(ZIEG/) ZIEGLER T E.	
XX			
PI		Deikman J, Feng PCC, Fincher KL, Ziegler TE;	
XX			
DR		WPI; 2004-479808/45.	
XX			
PT		New isolated nucleic acid molecule that encodes a plant protein or its	
PT		fragment, useful for isolating a variety of agronomically significant	
PT		genes associated with plant growth, quality or yield, and as molecular	
XX		tags to map genes.	
XX			

Db 1364 AAAAAAAAAATTTTATTAATAAAATTTT-----TAAATATTTTATTAATAAA 1330

Oy 305 TGTATTTAGCTTTATAGAAATGCTGTTATTTTGTAGAGCAGTGAAATTTATTAATTTT 364

Db 1329 TTATATATTAATTTTATTAATAATTTATTAATTTTAAAAAATATATACGTTTATATATTA 1270

Oy 365 TAGTACTGTTTCTTTTGGGTTCACTAGTACTCTTTTCTTGACATCAAAAT 424

Db 1269 ATATATATTCGTAACGTTTCTTTTATTCGAAAAAATATTTTATTAATAAAAAATTA 1210

Oy 425 ATTGTAGACGAGTGTCCATATATATAGATGTCGAAATGAAATGAAATATGACTTAATAATA 484

Db 1209 ATTAATAATATATTAATAATTTTAAAAATAATTAATATATATGAAAAATATATTAATTTTAA 1150

Oy 485 AATATAGAAATGTAACAAAAAATTTAGTGAAGAAAGTACTGAAAGAAAGCCTTA 542

Db 1149 ATTTTCGAATATCGAAAAAATATTTAAAAAAGCAATATATTAACGTTTAAAAACGTTAA 1092

RESULT 6

ABK39936

ID ABK39936 standard; DNA; 11422 BP.

XX

XX ABK39936;

DT 21-MAY-2002 (first entry)

XX

DE Human chemically pretreated gene sequence #9 strand 1.

XX

XX Human; day; bisulphite treatment; CpG; DNA methylation; cancer; tumour;

KM cytosine; ALDH6; CYP11A; CYP11B; CYP2A3; DPYD; EPHX2; OCLN; TXNRD1;

KM UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.

OS Homo sapiens.

XX

XX W020202806-A2.

PD 10-JAN-2002.

XX

XX 29-JUN-2001; 2001WO-EP007470.

XX

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX

XX (EPIG-) EPIGENOMICS AG.

PA

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-154757/20.

DR

XX

XX

XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,

PT useful for detecting cytosine methylation state of genes associated with

PT pharmacogenomics and for therapy of diseases e.g. cancer.

PS Claim 1; SEQ ID NO 17; 24bp; English.

XX

XX The invention relates to a nucleic acid comprising a sequence at least 18

CC bases in length of a segment of the chemically pretreated DNA of genes

CC associated with pharmacogenomics according to one of the sequences of the

CC genes ALDH6 (NM_000692), CYP11A (NM_000781), CYP11B (NM_000497), CYP3A3

CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN

CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,

CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and

CC their complementary sequences, or a sequence (S1) chosen from 87

CC sequences and their complements. The chemical pretreatment is bisulphite

CC treatment to convert cytosines (but not methyl-cytosines) into uracils.

CC Also included are an oligomer (II) in particular an oligonucleotide or a

CC peptide nucleic acid (PNA)-oligomer, comprising an oligonucleotide or a

CC base sequence having a length of 9 nucleotides which hybridises to or is

CC identical to a chemically pretreated DNA of genes associated with

CC pharmacogenomics and their complements, arranged in an array for

CC analysing diseases associated with the methylation state (CpG) and/or

CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The

CC	oligonucleotides may also be used as PCR primers. The set of 87 nucleic acids
CC	and their complements is useful for diagnosis and therapy of solid
CC	tumours and cancer. The present sequence represents one the 87 DNA
CC	sequences or its complement. Note: The sequence data for this patent did
CC	not form part of the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 11422 BP; 3867 A; 59 C; 1786 G; 5710 T; 0 U; 0 Other;
Query Match	8.6%; Score 60.6; DB 6; Length 11422;
Best Local Similarity	48.0%; Pred.No.0.097;
Matches	272; Conservative 0; Mismatches 284; Indels 11; Gaps 3
Oy	6 TTTAAATTGGGAAAAATTAATAATATATGTATTAAATAGCAAAATAAAGTTAGTGAGA 65
Db	7207 TTAAAATTAGTAGGAATATTAATAGTATTTTTTAAAAATGAATTAATAAATAAATAA 7266
Oy	66 GTTTTTTTAAATTTTTTATTAATAAATATTTTTGCATGAAATTTGTTTTAAGATAAA 125
Db	7667 GATTGTATTTTTTTTTTAAATGAAAGTATTTGTATTAATAGTATTAATTT---TTTAA 7322
Oy	126 TTTTGACATAACAAGATTTATTTTCGCCGCGCAGATGAGCTGAGAGATGTCCTGTC 185
Db	7323 AATGTGTAGTAAATTTTTTTTAAATTTAAATTAATTAATTTAGTTTTTAAATTTAAATTT 7382
Oy	186 TTTAAACCTCGTAAAGTTTGGTCTTACCACCAATVCCACAAAGTAAACGACCAT 245
Db	7283 TTTAAAATTAAATTTTTTAAATTTGAATAGTTGAAATTTTAAAT---AAAAATTTT 7438
Oy	246 ACCGGTTATATATTTCTAACCGGTTTAAAGTTTACATTAATCATTTACTTAATCCGGT 305
Db	7439 TGTTTTTATTTTAAATTTAAATTTTATTTGATAGTAAATTTGATTTAAATTTTAGATAT 7498
Oy	306 GTAATTAAGTTTATNGAAATGCGTTATTTTGTAGTGCAGGTGAATTTATTAATTTT 365
Db	7499 ATAGTGTATTTTGAAAAATGTTTTTATTTTATGATAATGTTTAAATGTTTATGTTTT 7558
Oy	366 AGTACTGTCTTTCTTTTGGGTTCAACTAGTACTCTTTTCCTTGACATCAAAATTA 425
Db	7559 TTTGAATTTATGAGATAGTATTTAATTTGTATTTAGAAATTT---ATTAAAGAAAACGA 7615
Oy	426 TTGTAGACGAGTGTCATATATATGATGCGAAATGAAATGATATTTAGTATTAATTA 485
Db	7616 TTTAAATTTGACGTGTGTATATATATTTTAAATTTAAAGTATTAATTAATTAATTA 7675
Oy	486 ATATAGAAAGGAAACAAAAAAATTAATGAGAAAAGTAACTCGAAAGAAAAAGCTTAT 545
Db	7676 AAAAAAAAAAAGAAACGAATATGTTTTGTGTAATTTTAAATTTAATAAATTTTGAGA 7735
Oy	546 GCTTGCAAGATTCATAGTTTATTT 572
Db	7736 TATGATTTAATATTTTATTTTATTT 7762
RESULT 7	
ABL32218	
ID	ABL32218 standard; DNA; 11422 BP.
XX	
AC	ABL32218;
XX	
XX	26-MAR-2002 (first entry)
DE	Human immune system associated gene SEQ ID NO: 191.
KW	Human; immune system disease; cytosine methylation; antiasthmatic;
KW	antiartherosclerotic; antianaemic; cytostatic; neoplastic;
KW	neuroprotective; anti-HIV; anticovulsant; ophthalmological;
KW	antiinflammatory; anticholelithic; antidiabetic; antipsoriatic;
KW	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW	acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW	neutrofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW	ds.

[illegible][illegible]

CC Present sequence represents a specifically claimed EST isolated from a
CC cotton variety DP50B primed seed cDNA library (LIB3825). The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX

SO Sequence 560 BP; 268 A; 7 C; 48 G; 237 T; 0 U; 0 Other;

Query Match 8.6%; Score 60.4; DB 13; Length 560;
Best Local Similarity 50.7%; Pred. No. 0.097;
Matches 145; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
QY 251 TTAATTAATTTCTACCGGTTTATTAAGTTTACTTAATCAATTACTTAATCCGCTTAAT 310
DB 371 TTTTCTTTTAAACACCTTTTCTTTTAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTT 312
QY 311 TAAAGTTTAAAGTGTGTTTATTTAGTGTGACGTGAATTTATTTATTTTGTATC 370
DB 311 TTTTCTTTTAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 252
QY 371 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 430
DB 251 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 192
QY 431 GACGAGTGTCCATATATAGATGTGAAATGAATGATATGTAGTAAATTAATATATA 490
DB 191 AAATTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 132
QY 491 GAAAGTGAACAAAAAATTAAGTGAAGAAAGTAACTGGAAGAAAGAA 536
DB 131 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 86
```

RESULT 9

ID ABV60941/C
ABV60941 standard; cDNA; 626 BP.

AC ABV60941;

DT 13-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 60932.

KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US005171.

PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0213314P.

PR 18-JUL-2000; 2000US-0215007P.

PR 13-DEC-2000; 2000US-0255281P.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JB;

DR WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1; Page 11585; 11750pp; English.
XX

CC The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

SO Sequence 626 BP; 356 A; 15 C; 41 G; 130 T; 0 U; 84 Other;

Query Match 8.5%; Score 60.2; DB 5; Length 626;
Best Local Similarity 48.0%; Pred. No. 0.11;
Matches 134; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```
QY 258 TATCTAACCGGTTTAAAGTTTAAATCAATTAATCAATTAATTAATTAAGTTT 317
DB 330 TNNCCNCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 271
QY 318 TATGAAGTGTGTTATTTGTAGTGTGACGTGAATTTATTAATTTTGTAGTCTGTTT 377
DB 270 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 211
QY 378 TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 437
DB 210 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 151
QY 438 GGTCCATATATAGATGTGAAATGAATGATATGATTAATTAATTAATTAAGTGTG 497
DB 150 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 91
QY 498 AACAAAAAATTAAGTGAAGAAAGTAACTGGAAGAAAGAA 536
DB 90 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 52
```

RESULT 10

ID ACN46935
ACN46935 standard; cDNA; 499 BP.

AC ACN46935;

DT 02-DEC-2004 (first entry)

DE Cotton primed seed EST Clone ID: LIB3825-005-Q1-N6-D10, SEQ:1716.

XX

KM Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;

KW variety DP50B; library LIB3825; molecular tag; molecular marker;

KW genetic mapping; molecular mapping; seed germination; plant growth;

KW plant quality; plant yield; plant breeding; tissue printing; ss.

OS Gossypium hirsutum.

PN US2004123340-A1.

PD 24-JUN-2004.

PF 12-DEC-2001; 2001US-00021323.

PR 14-DEC-2000; 2000US-0255619P.

PA (DEIK/) DEIKMAN J.

PA (FENG/) FENG P C C.

PA (FING/) FINCHER K L.

PA (ZIEG/) ZIEGLER T E.

XX Delkman J, Feng PCC, Fincher KL, Ziegler TE;

XX WPI; 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its
PT fragment, useful for isolating a variety of agronomically significant
PT genes associated with plant growth, quality or yield, and as molecular
PT tags to map genes.

XX Claim 1; SEQ ID NO 1716; 34pp; English.

XX The invention relates to 17880 cotton expressed sequence tags (ESTs);
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC from primed or non-primed seeds from variety DP50B, mature seeds from
CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC tissue, developing fibres, carpel walls and septa from variety
CC Nuotcon33B. The invention also relates to substantially purified
CC the proteins or their fragments encoded by nucleic acid molecules of the
CC invention, and to transformed plants having a nucleic acid construct
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC molecular tags to isolate genetic regions, to isolate genes, to map
CC genes, to determine gene function and to determining whether genes are
CC members of a particular gene family. The nucleic acid molecules may be
CC used for isolating a variety of agronomically significant genes
CC associated with plant growth, quality, yield, and could also serve as
CC links in metabolic and catabolic pathways. The nucleic acid molecules are
CC also useful for identifying genes important in initiating and maintaining
CC seed germination or that may be used to mitigate stresses encountered
CC during seed germination. The ESTs additionally enable the acquisition of
CC promoters and cis-regulatory elements which will be useful to express
CC agronomically significant genes in these tissues and/or other tissues,
CC and also permits the acquisition of molecular markers useful in breeding
CC schemes, genetic and molecular mapping, and in cloning of agronomically
CC significant genes. The nucleic acid molecules are further useful for
CC detecting the presence or quantity of a protein by tissue printing. The
CC present sequence represents a specifically claimed EST isolated from a
CC cotton variety DP50B primed seed cDNA library (U183825). The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?docID=US20040123340
XX

XX Sequence 499 BP; 212 A; 32 C; 62 G; 193 T; 0 U; 0 Other;

XX Query Match 8.5%; Score 60; DB 13; Length 499;

XX Best Local Similarity 50.7%; Pred. No. 0.11;
Matches 144; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

XX 305 TGTATTAAGTTTGAAGTGGTATTTGTAGTACGACGAAATTTATTAATTTT 364

XX 34 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 93

XX 365 TAGTACTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 424

XX 94 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 153

XX 425 ATTGTAGACGAGTGTCCATATATAGATGATGAAATGAATATTTAGATTAATAA 484

XX 154 AAAAAAGGGGGGTAAAAAATTTTAAAAAGAAAAAAGAGGGGGGTAAAAA 213

XX 485 AATTATGAAGATGACAAAAAATTAAGTGAAGAAAGTAACTGGAAGAAAGCTTAT 544

XX 214 AAGGGGGGAAAAA 273

XX 545 GCGTGGCAAGATTCCATTAAGTTTATTCACCAAAAGGAAAGA 588

XX 274 AAAAAAGGAAAAAAGGGGTTTAAAAAAGGGGGGAAAAA 317

XX RESULT 11

XX ID ABZ10100 standard; DNA: 8056 BP.

XX AC ABZ10100;

XX DT 16-JAN-2003 (first entry)

XX Haematopoietic cell proliferation disorder related DNA sequence #240.

XX Human; haematopoietic cell proliferation disorder; cytostatic;

XX gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;

XX cytosine methylation state; gene; ds.

XX Homo sapiens.

XX WO200277272-A2.

XX 03-OCT-2002.

XX 26-MAR-2002; 2002WO-EP003401.

XX 26-MAR-2001; 2001US-027833P.

XX (EPIG-) EPIGENOMICS AG.

XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
XX Olek A, Piepenbrock C, Adorjan P, Grabs G, Iesche R, Ieu E,
XX Lewin A, Lipscher B, Walter S, Model F, Mueller V, Otto T, Pelet C;
XX Schwabe I, Ziebarth H;

XX WPI; 2003-018942/01.

XX Detecting and differentiating between haematopoietic cell proliferative
XX disorders, comprises contacting a target nucleic acid with a reagent that
XX distinguishes between methylated and non-methylated CpG dinucleotides.

XX Claim 28; SEQ ID NO 240; 117pp; English.

XX The present invention describes a method for detecting and
XX differentiating between haematopoietic cell proliferative disorders
XX associated with at least 1 gene and/or their regulatory regions in a
XX subject. The method comprises contacting a target nucleic acid in a
XX biological sample obtained from the subject with at least 1 reagent,
XX which distinguishes between methylated and non-methylated CpG
XX dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118

XX represent specifically claimed nucleotide sequences from the present
XX invention. Oligonucleotides from the present invention can be used: for
XX differentiating between healthy haematopoietic cells and proliferative
XX disorder haematopoietic cells; for differentiating between acute
XX lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
XX determining the cytosine methylation state and/or single nucleotide
XX polymorphisms (SNPs) of haematopoietic cell proliferation disorder
XX related sequences and their complements; and as primers for the
XX amplification of haematopoietic cell proliferation disorder related DNA
XX sequences. The nucleotide sequences from the present invention can also
XX be used for detecting a predisposition to, differentiation between
XX subclases, diagnosis, prognosis, treatment and/or monitoring of
XX haematopoietic cell proliferative disorders. The present method enables a
XX highly specific classification of haematopoietic cell proliferative
XX disorders allowing for improved and informed treatment of patients

XX Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

XX Query Match 8.4%; Score 59.4; DB 8; Length 8056;

XX Best Local Similarity 44.7%; Pred. No. 0.16;
Matches 272; Conservative 0; Mismatches 336; Indels 1; Gaps 1;

XX 11 ATTGGGAAAAATTAAT 70

XX 2109 ATTAT 2168

XX 71 TTTTAAATTTTAT 130

XX 2169 ATTAT 2228

XX 131 ACATTAACCAAGAT 190

XX 2229 AATTTAAT 2288

PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

XX WPI: 2002-154758/20.

XX Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signalling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signalling.

XX Claim 1; SEQ ID NO 358; 24pp + Sequence Listing; English.

XX The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligonucleotides for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records AB170111-AB170626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office

XX Sequence 15954 BP; 4759 A; 139 C; 3187 G; 7869 T; 0 U; 0 Other;

Query Match 8.4%; Score 59.4; DB 6; Length 15954;
Best Local Similarity 51.4%; Pred. No. 0.16;
Matches 162; Conservative 0; Mismatches 151; Indels 2; Gaps 1;

OY 230 AAGGTAACGACGACCGTATTAATATCTTACCGTTTAAAGTTTACATAAATC 289

Db 11745 AATTAATAATGTTTATTAAGTTAATTAATTAATTAATTAATTAATTAAT 11804

OY 290 ATTACTAATCCGCGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 349

Db 11805 ATTGATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11864

OY 350 AATTATTAATTTTAACTGTTGTTTCTTTTGGGTTCACTAGTACTTTTTC 409

Db 11865 AATGATTTGATTTTATTTTATTTTGTAAATGATTAATTAATTAATTAATTAAT 11924

OY 410 TTTCATCAATAATTTTGAACGAGTGTCCATATAGATGTAATGAATGAAT 469

Db 11925 TTGAGATT--AATGAATCGAATTAATTAAGGAGTATTTTACGAGAGAGATTAAT 11982

OY 470 ATTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 529

Db 11983 TTAGAAAAATTAAGAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 12042

OY 530 AAGAAAAAGCTTAT 544

Db 12043 TAATATTATTATT 12057

RESULT 14

ADQ24856/c

ID ADQ24856 standard; DNA; 6816 BP.

XX ADQ24856;

XX 26-ANG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 7676.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KM de.

OS Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

XX 26-NOV-2002; 2002US-0429739P.

XX (PROT-) PROTEIN DESIGN LABS INC.

XX Aziz N, Ginsburg WM, Zlocnik A;

XX WPI: 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.

XX Example 2; SEQ ID NO 7676; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 6816 BP; 2121 A; 1327 C; 1578 G; 1571 T; 0 U; 219 Other;

Query Match 8.4%; Score 59.2; DB 12; Length 6816;
Best Local Similarity 44.3%; Pred. No. 0.17;
Matches 186; Conservative 0; Mismatches 222; Indels 12; Gaps 2;

OY 5 TTTTAAATTTGGGAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 64

Db 6653 TTTATATNNNANTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6594

OY 65 AGTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 124

Db 6593 TTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6534

OY 125 ATTTGACATTAACCAATTAATTTTCCGCCAGATTAAGTCTAGAGAGATGCTCGT 184

Db 6533 TTTTATTTTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6474

OY 185 CTTTAACTCGTAAGTTTGTCTTACCAACCAATACCAATTAAGTAACGACCA 244

Db 6473 AAAAAAANTTAATAATTTTATTTTATTTT-----TTNCNNAAAAAATAATATAA 6419

OY 245 TACCGTATTAATTAATTTTCAACCGTTTAAGATTTACATAATCTTACTAATCCGG 304

Db 6418 NNNANTTAATAATTTTATTAATAAATNNNTTTTAATAAATAATTTTATTTT----- 6364

OY 305 TGTAAATTAAGTTTATGAATGCTGTTATTTGTAGTCAACGTAATTAATTAATTTT 364

Db 6363 --TAAAAAATAAATAATTTNNNNNTTTTATTAATTAATTAATTAATTAATTAATTAAT 6306

OY 365 TAGTACTGTTTCTTTTGGGTTCACTGTAATTTTCTTTCTTGACATCAAAAT 424

Db 6305 NTTTTNNNNNNNNNTTTTNTTTNNNNNTTTTNTTTTNTTTNNNNNTTTNTTT 6246

RESULT 15

ABLS5644
 ID ABL5644 standard; DNA; 50000 BP.
 XX
 AC ABL5644;
 XX
 DT 01-JUL-2002 (first entry)
 XX
 DE AMEPV genome fragment#2.
 XX
 KM AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
 XX genetic deficiency disorder; ds.
 OS Amsacta moorei entomopoxvirus.
 XX
 PN WO200212526-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 10-AUG-2001; 2001WO-US025287.
 XX
 PR 10-AUG-2000; 2000US-0224479P.
 PR 14-SEP-2000; 2000US-00662254.
 XX
 PA (UYFL) UNTV FLORIDA.
 XX
 PI Moyer RW, Li Y, Bawden AL;
 XX
 DR WPI; 2002-227161/28.
 XX
 PT Novel recombinant entomopox virus vector useful for delivering
 PT polynucleotide encoding protein to vertebrate cell, comprises
 PT polynucleotide encoding protein operably linked with heterologous
 PT promoter sequence.
 XX
 PS Disclosure; Page 150-175; 326pp; English.
 XX
 XX The invention relates to a recombinant entomopox virus (EPV) vector,
 CC comprising a polynucleotide encoding a protein operably linked with a
 CC heterologous promoter sequence. The invention also concerns methods for
 CC providing gene therapy for genetic deficiency disorders. Vectors of the
 CC invention are useful for delivering a polynucleotide encoding a protein
 CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
 CC The vector is introduced into the vertebrate cell by infection in a viral
 CC particle, or by transfection, transduction, or injection either in vitro
 CC or in vivo. The vector is useful for the delivery and expression of
 CC biologically useful proteins in gene therapy protocols, and for
 CC delivering large DNA segments for engineering of vertebrate cells.
 CC Polynucleotides of the invention have applications in techniques such as
 CC their use as insertion sites for foreign genes of interest, hybridisation
 CC probes, for chromosome and gene mapping, in PCR technologies, and in the
 CC production of sense or antisense nucleic acids. Vectors of the invention
 CC provide for stable integration and expression of heterologous DNA in host
 CC cells, and are adapted for accepting large heterologous polynucleotide
 CC inserts which can be delivered in an infected or transformed cell and
 CC expressed in a stable fraction. The current sequence represents a
 CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei
 CC (AMEPV)
 CC
 SQ Sequence 50000 BP; 20514 A; 4505 C; 4614 G; 20367 T; 0 U; 0 Other;
 Query Match 8.4%; Score 59.2; DB 6; Length 50000;
 Best Local Similarity 54.8%; Pred. No. 0.18;
 Matches 138; Conservative 0; Mismatches 113; Indels 1; Gaps 1;
 QY 260 TTCTAACCGGTTTATAGTTTACATTAATCATTTACTAATCGCGTAAATTAAGTTTA 319
 DB 49095 TTATTAATCTTATTTTATTAATTTATCAATCATTTTATATATATATTATTCAAATTAAT 49154
 QY 320 TGAATAAGGTTATTTTGTAGCGACGTAATAATTTATTAATTTTGTAGTCTGTTTC 379
 DB 49155 TAAATATTTTATTAATTTTCTATATATAGATATTTTATTAATTTTGTAGTATTAAT 49214
 QY 380 TTTTGGGTCAACTAGTACTTTTCTTGTGACATCAAAATTAATTTGTAGACGAGTGG 439

DB 49215 AAATTTATTAATTAATATATATATATATATGTTTAAATTTTAAATCAATTATTTTAAACATGCCA 49274
 QY 440 TCCATATATAGATGCGTGAATGAAATGCAATATATGAGTAAATTAATATAGAAAGTGA 499
 DB 49275 AAATTAATATATATATGTAATATATTTTATTTATA-GATTTTTCATTAATAATTAATTTGTT 49333
 QY 500 CAAAAAAATTA 511
 DB 49334 CTAAATATATTA 49345

Search completed: November 17, 2005, 15:59:37
 Job time : 479 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2005, 14:59:33 ; Search time 3117 Seconds

(without alignments)
8609.339 Million cell updates/sec

Title: US-10-810-788a-4

Perfect score: 705
Sequence: 1 aagcttcaattcgggaaa.....ctctggtcacttgatacc 705

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	75	10.6	1225	9	CNS0161D
C 2	74.4	10.6	1101	9	AL069706 Drosophila
C 3	74	10.5	1101	9	CNS00807
C 4	72.4	10.3	524	9	CNS01090
C 5	72.4	10.3	1101	9	CNS0039G
C 6	70.8	10.0	734	9	CNS0160P
C 7	68.8	9.8	1200	9	CNS0106CO
C 8	67.8	9.6	1896	3	CR722884
C 9	67.2	9.5	928	9	CNS00DKY
C 10	67.2	9.5	1101	9	CNS0021J
C 11	67	9.5	1101	9	CNS00EVL
C 12	66.8	9.5	886	8	BH177277
C 13	66.8	9.5	886	9	CNS07JUX
C 14	66.6	9.4	658	9	CNS03NXX
C 15	66.4	9.4	1592	9	CG750135
C 16	66.2	9.4	625	9	CNS036A2
C 17	66.2	9.4	928	9	CNS00DKY
C 18	65.8	9.3	782	7	CV484987
C 19	65.8	9.3	1101	9	CNS0038B
C 20	65.6	9.3	826	9	AG579614
C 21	65.4	9.3	1101	9	CNS001F8
C 22	65.4	9.3	1101	9	CNS003BD
C 23	65.4	9.3	1147	6	CD387645
C 24	65	9.2	1200	9	CNS016CO
					AL106578 Drosophila

25	64.8	9.2	1168	9	CL078758
C 26	64.2	9.1	1001	9	CNS01400
C 27	64.2	9.1	1043	9	CNS0145P
C 28	64.2	9.1	1542	9	AG386981
C 29	64	9.1	1156	9	AG396050
C 30	63.8	9.0	940	9	CNS00LDR
C 31	63.6	9.0	1201	9	CNS0167M
C 32	63.2	9.0	1101	9	CNS0039G
C 33	63	8.9	625	9	CNS036A2
C 34	63	8.9	888	8	AZ549422
C 35	63	8.9	1201	9	CNS0152Z
C 36	62.2	8.8	752	7	CF519392
C 37	62.2	8.8	888	7	CK156432
C 38	62	8.8	709	7	CV480279
C 39	62	8.8	1210	9	CG749728
C 40	61.8	8.8	646	7	CF547155
C 41	61.8	8.8	836	9	CNS01100
C 42	61.8	8.8	987	9	CNS014PO
C 43	61.8	8.8	1101	9	CNS00EPO
C 44	61.8	8.8	2077	3	CR67281
C 45	61.6	8.7	1190	9	CNS020N7

ALIGNMENTS

RESULT 1

LOCUS CNS0161D 1225 bp DNA linear GSS 26-JUL-1999

DEFINITION Drosophila melanogaster genome survey sequence Sp6 end of BAC
BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL106171.1 GI:5620504

KEYWORDS GSS.
Drosophila melanogaster (fruit fly)

SOURCE Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1225)
Genoscope.
Direct Submision
Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT Determiation of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES

source 1..1225
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_1ib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : Sp6"

ORIGIN
Query Match 10.6%; Score 75; DB 9; Length 1225;
Best Local Similarity 29.0%; Pred. No. 0.0013;
Matches 165; Conservative 129; Mismatches 274; Indels 0; Gaps 0;

OY 5 TTTTAAATTTGGGAAATTTAATATATATGTTAATGCAATTAAGATTAGATGAG 64
DB 1204 TTTTAAATTTGGGAAATTTAATATATATGTTAATGCAATTAAGATTAGATGAG 64

	Db	324	AATAAATAAAMATTAAAAATTTTTTTTWTTEATWTTTAATTTTTTAAATTTTTTAAATTTTTTAAAT	383
	Oy	301	CGGCTGTAAATTAGTTTATGAACAATTCGGTATTTTGAGGTACACGTAAATTATTAT	360
	Dd	384	WATMTTATATTTTTTTTTTATATTTTTTATTTTTTATTTTTTATTTTTTTTTTTT	443
	Oy	361	TTTTTAGACTTGTTTTCTTTTTGGGTCAACTAGTTACTTTTCCTT	412
	Dd	444	TT	495
<hr/>				
RESULT 5 CNS0039G/c CNS0039G/c LOCUS				
DEFINITION Drosophila melanogaster genome survey sequence TETJ end of BAC # BACR08K10 of RPcI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION AL063921				
VERSION AL063921.1 GI:4941778 GSS.				
SOURCE	ORGANISM	Drosophila melanogaster (fruit fly)		
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryoeta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	AUTHORS	Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. for further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosses at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPcI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.		
FEATURES	SOURCE	Location/Qualifiers 1..1101 /organism="Drosophila melanogaster" /mol_type="genomic DNA" /db_xref="taxon:7227" /clone="BACR08K10" /clone_1fb="RPcI-98" /note="end = TETJ"		
<hr/>				
Query Match	10.3%; Score 72.4; DB 9; Length 1101;			
Best Local Similarity 18.5%; Pred. No.0.0039; Mismatches 264; Indels 2; Gaps 2;				
Matches 125; Conservative 283;				
Oy	30	TATATGTATTAAATGCACAATAAAAAGTTAGATGACAGTTTTTTAAATTTTTTATATATA	89	
Dd	1100	KARWBGDDTWDRDJRKODMDWKMTWMKKDBADDERMAVGDAADRMAWDGAGTMMTATMWW	1041	
Oy	90	AATAATTTTTTGCAGAATTGTTTTTAAGATAAATTTTGAACATAACCAAAGTATATTT	149	
Dd	1040	WWWAWTDTDMDKMWMMWATAKTDTATMTWRTRAFRADWAGRDGARGRRDRAATDADGA	981	
Oy	150	TCCGCCACAGATTTGAGTCTACGAGAGATGCTCTGCTTTAAACCCTGAAAGTTTGATC	209	
Dd	980	GRRDGGRKRKDCKORGDSDIKGGGKKKAALKAKMATKMWDMDWDKOMTKWGAKDKKA	921	

[illegible]

RESULT 7	
CNS016CO	
LOCUS	CNS016CO 1200 bp DNA linear GSS: 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence 17 end of BAC BACN15E04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL106578
VERSION	AL106578.1 GI:5622626
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyraidae; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1200)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 101 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a

QY		191	ACGTCGAAAGTTTGGCTTACCACAACCAATACCAACAAGGTAAACGCATACCG	250
Db		761	KHYYCSCSMWTATTCYMAIYVCSGRHAAMMHCCSCSCHAIAAAAAAAAAMMMMMMGGG	820
QY		251	TTAATAATATTCTPACCGGTTTTAAGTTTACATTAATCATTAATCGCGCTGAAT	310
Db		821	KTRRRCGYCTDCITMAAAGTTTTTHGGTTWMMWATAMWAIAAAAAAAAAAATTTAAAA	880
QY		311	TAAGTTTATNGAANAIGGTTATTTTGTAAGTCACGGAAATTTATTAATTTTAGTAC	370
Db		881	TTTTYTAIAAAAAATPADGAWTAATTTAACGAAAAATTTTAAWATTTATTTWWAAAIAAW	940
QY		371	TTC-TTTTTCTTTTTGGGTCAACTGACTCTTTTTCCTTGACATCAAAATTAATGT	429
Db		941	AABAATWTTTTTAAATTAATWMTWYTWTWTAATTTTAAAMWAIAAAATTTTTTTTWAAAT	1000
QY		430	AGACGAGTGTCCATATATGATGGTGCAGAAATGAATGAATATTTGAGTAATTAATAATAT	489
Db		1001	AAMWTAMAWMATWMAAAWMTATTTAAATATTAIAAAATTAWATWMAWMAAAWMAIAAAAA	1060
QY		490	AGAAATGAAACAIAAAAAAATTTAGTGGAAAGATPACTCGAAGAAGAAAGCTTATGGCTT	549
Db		1061	AAMWATWMAAABAAAAAABAAAAAABAAAAAABAAWTTAATTAIAWMAWMAATTAWMAATTT	1120
QY		550	GCGAAGTTCATPAAGTTATTTTACCAAIAAGAAAGAAAGTACTGGCTTGCTCTCT	609
Db		1121	WTAAAAAATWATAWATWATTTTTTTWMTWMAATTTAATTTATWMTWMAATTAWATAWAT	1180
QY		610	CTCTTTTAAAT 621	
Db		1181	ATATATWATDAW 1192	

RESULT 8					
CR722884	CR722884	1896 bp	mRNA	linear	HTC 19-AUG-2004
LOCUS	Tetraodon nigroviridis full-length cDNA.				
DEFINITION	CR722884				
ACCESSION	CR722884.1	GI:51221135			
VERSION	HTC; cDNA; full-length				
KEYWORDS	Tetraodon nigroviridis				
SOURCE	Tetraodon nigroviridis				
ORGANISM	Tetraodon nigroviridis				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.				
REFERENCE	1 (bases 1 to 1896)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-AUG-2004)	Genoscope - Centre National de Sequencage -			
	2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE				

COMMENT					
	(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
	The sequences are based on single pass reads.				
	More information available at http://www.genoscope.cns.fr/tetradon. location/Qualifiers				
FEATURES					
source	1..1896 /organism="Tetradodon nigroviridis" /mol_type="mRNA" /db_xref="taxon:99883" /cisuse_type="Kidney"				
ORIGIN					
Query Match	9.6%;	Score 67.8;	DB 3;	Length 1896;	
Best Local Similarity	47.6%;	Pred. No. 0.024;	Mismatches 269;	Conservative 0;	Indels 9; Gaps 2;
Oy	5	TTTTAAATTTGGGAAAATTATATATATATGTATTAATGCACAATAAAGTAGATGAC	64		
Db	123	TTTTTAATTTAAATTTTAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATA	182		
Oy	65	AGTTTTTTTTAAATTTTTTATATATATTAATTTTTTGTCATGAATTTGTTTAAAGATAA	124		
Db	183	AAATTTAAAAATTTTAAAAAAAATTAATTTATTTTAAATTTTAAATTTTAAATTTAA	242		
Oy	125	ATTTTGACATPACCAGATGATTTTCCGCCAGCAATTTGAGCTCAGACGAGTGCTGT	184		
Db	243	A-----AAAAAAAAAAAAATTTTAAAAAAAATTAATTTTAAAAAAAATTTTAAT	295		
Oy	185	CTTTAAACCTCGTAAAGTTTGTCCTAACCCACCAATCCCAAGAAGTAAAGCACCA	244		
Db	296	AAAAAAAAAAAAAAAAAAAAAAAAATTTAAATAATTTAAATTAATTTTAAAAAAAATTTAA	355		
Oy	245	TACCGTAAATATATYCTTAAACGGTTTAAAGTTTACATPAATCATTTACTAATCCGCG	304		
Db	356	AAAATTTAAAAAATTAATAAATAATTTATTAATAAATAATTTAAATAA--AAAA	413		
Oy	305	TGTAATTAAGTTTATGAATGCTGTATTTTGACGTCAGTGAAATTTTAAATTTT	364		
Db	414	TTTTATTAATTAATTTTATTTTATTAATAATTAATAAATAAATAAATAATTAATTTT	473		
Oy	365	TAGTACTGTTTTCTTTTGGGTCACATGATTTACTTTTTCTCTTGACATCAAAAT	424		
Db	474	AAATATTAATTTATATATATTTTATTTTATTAATTAATTAATAAATAAATAAATTAAT	533		
Oy	425	ATTGTAGACGAGTGTCCATATATAGATGTGAATGAATGAATGATTTGAGTAATTAATA	484		
Db	534	ATTTTAAATATTTTATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA	593		
Oy	485	AATATAGAAAGTGAACAAAAAATAATAGTGAATAAGTAACTGGAAGAAGAAAGCTATT	544		
Db	594	TTTTTGTAAAGTTATTAATAAATAATTTGTTAAATTAATTAATTAATTAATTAATTA	653		
Oy	545	GCGTTGGCAAGATTCATTAAGTTTA	569		
Db	654	TTAATTTTATTAATTAACCTATTTTA	678		

RESULT 9
CNS00DKY/C 928 bp DNA linear GSS 04-JUN-1999

LOCUS DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
BACH27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL071865
VERSION AL071865.1 GI:4948170

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 928)
Genoscope.

TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT

- Web : www.genoscope.cns.fr

determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP Drosophila melanogaster BAC library was prepared by Kazuo Osoegawa and Aaron Mammeter in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPcI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source location/Qualifiers

1..928

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone="BACR27A24"

/clone_lib="RPcI-98"

/note="end : T7"

ORIGIN

Query Match 9.5%; Score 67.2; DB 9; Length 928;

Best Local Similarity 30.7%; Pred. No. 0.03;

Matches 99; Conservative 86; Mismatches 137; Indels 0; Gaps 0;

OY 198 AAAGTTTGGCTGTACCACAACCAATGCCAACAAAGTAACGACTACCGGTTAATA 257

Dd 864 AAAAMWTCCAAAAMAMAMCMVMMMMMMMMTTTTMMNNNNNNHHHMTTTT 805

OY 258 TATTCTAACCGGTTTAATAGTTTACATAAATCATTTACTAATCCSGGTGAATTAAGTT 317

Dd 804 TTWTTTTTTTTTTTTTMMMMMMWWTTTTMMWTTTTMMNNNNCHMAATMW 745

OY 318 TATGAAATGTGGTATTGTTAGSTCAGTGAAATTATTAATTTTTTAGTACTGTTT 377

Dd 744 MMWAATTTWTAATWTAAMATWAAAAAAMAAAAAATTTTWTWTTAATWTA 685

OY 378 TCCTTTTGGGTCACTAGTACTCTTTTCCTTGACATCAAATATTATGTAGACAGT 437

Dd 684 AATTTTTTATTTTTTTTTTTTTTTTWTATATAAATWTAAMAAAMATATTATATW 625

OY 438 GGTCATATATAGATGCGTAATGAATGAATATTGATGATATAATAATATAGAAGT 497

Dd 624 ATATTWTAATWTAATATAAATWTAAMWTTTAATWTAAMATWAAATWTAATAA 565

OY 498 AACAAAAAAATTAGTGAAA 519

Dd 564 AAAAAAATAATWTTWTA 543

RESULT 10 CDS0021J/c 1101 bp DNA linear GSS 03-JUN-1999

LOCUS CDS0021J melanogaster genome survey sequence TET3 end of BAC #

DEFINITION BARO5N11 of RPcI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL061936 GI:4940214

VERSION AL061936

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tübingen D-72076, Germany
Tel.: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@uebingen.mpg.de
Class: BAC ends.

FEATURES
source

vector. "The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."

Search completed: November 17, 2005, 17:46:28
Job time : 3124 secs

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 FILE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knodde Martens Olsson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelson, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CPI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-0176
 TELEFAX: (619) 235-8550
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19124 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cdna
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 US-08-487-826B-13

Query Match	7.3k; Score 51.6; DB 2; Length 19124;
Best Local Similarity	46.1k; Pred. No. 0.13;
Matches 213; Conservative	0; Mismatches 244; Indels 5; Gaps 1;
Qy 67	TTTTTTAAATTTTTTATTTATTTAAATATTTTGGATGTAATTTGTTTAAAGATTAAT 126
Db 15781	TTTTTAACATTTTTTTAAATTTTTTTATTTTATTTTATGATATATATTTTATTTAAATATTT 157222
Qy 127	TTTGACATTAACCAAGTATTAATTTTCGCCACGAATTGAGCTACGAGAGTCTGTCT 186
Db 15721	TTTTCTTTTATTTTTCGTTTATGATATATTTTATTTTAAATGTTTTTTTTTTT 156622
Qy 187	TTAAACCTCGTAAAGTTTGGCTCTTACCCAAACCAATACCCCAAGTAAAGTAAAGACATTA 246
Db 15661	CTTCTTTTGTATTTTATTTTATTAATCATTTTTTTTTTATATATAATTTTTTTTTAT 156022
Qy 247	CCGGTATATATATTTCTAACCGGTTTATAGTTTACATTAATCATTTACTTAATCCGCGTG 306
Db 15601	TTTTTTTGTAAATCTTTTTCATTTTATTTTATCTATCAAAATTTATTTTATATTAATTT 155422
Qy 307	TAATTAAGTTTAAAGAAATGT-----GGTATTTTGTAGTCAAGTAAATTTATTAAT 361
Db 15541	TTATTAATTTTAAAAAAATTTTCTCCTTTTTTTTTTTTTTTTTTAAATTAATTAAT 154822
Qy 362	TTTTTAGTACTGTTTTTCTTTTTGGGTGCACTAGTACTTTTTTCTTTGACATGAA 421
Db 15481	TTTTTTTATATTTTCATTTTCTTTTCTTTTTCATTTTAAATATTTGTTTTTAATTTCTTT 154222
Qy 422	ATTATTTGTACGAGTGGTCCATTAATATAGATGGTGAATGAATGAATATGTAGATTA 481
Db 15421	TTTAAATTTAAATACATTAATATATTAATAATATATATATTAACATATCAATGATCATAT 153622
Qy 482	ATAAATATAGAAAGTGAACAAAAAATTTGTGGAAAAAGGTA 523
Db 15361	ATATATATCTTAAACCAAAATATATATACAGAAATTCATA 15320

```

US-09-949-016-12776/c
; Sequence 12776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/7241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 12776
; LENGTH: 187169
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(187169)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12776

```

[illegible]

RESULT 4
US-09-949-016-15940/c


```

;   FEATURE:
;   NAME/KEY: promoter
;   LOCATION: (1) ..(1141)
;   OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

```

Query Match 7.1%; Score 50; DB 4; Length 1141;
 Best Local Similarity 11.6%; Pred. No. 0.18; Indels
 Matches 74; Conservative 226; Mismatches 336; Gaps 1.

Qy	49	ATAAAGTAGAAGAGGTTTTTAAATTTTTTAAATTTTAAATTTTGAATGATAA	108
Db	936	WTTDVERMAKAKNNNNNNAAYTACNPAATNNKATMTMMKTHGASKRTTHHTCR	877
Qy	109	TGTGTTTTAAGATAAATTTTGCACATPAACCAAGTATATTTTCCGCACGATTAAGTCT	168
Db	876	RTKYNNNNNNNARIVYUHHAAARRMMAMWTRTNNNNNNNNNNNACRNTETWABKHSWCN	817
Qy	169	ACGAGAGATGTCCTGCTTTAAACCTCGTAAGTTTGCTGTACCCCAACCCATACCA	228
Db	816	NNNNNNNNNNNTWCHYTTTAAABBCRYANNNNNAARATCNNYMAAAVTTHTDVCYK	757
Qy	229	CAAGGTAAACGACCAATCCGGTTATATATTTCAACCGGTTATAGTATACATAAAT	288
Db	756	TMMNTWYMDMTTMBTTTTTRNMTTSNTNNNNNNNNMACTNNNNNNNKAYAHATNNWG	697
Qy	289	GATTTACTAATCCCGCTGTAATTAAGTTTATGAATGCTGTAATTTGTAGTCACTG	348
Db	696	CMNNNTDARITNNTTVMRRMMNTNKTIRYSTRRHHTYGATNNNNNNNNNNNNNSC	637
Qy	349	AAATTTAATTAATTTTGTAGTCTGTTTTCTTTTGGG---TTCAGTAGTACTTT	404
Db	636	CTCTRMATMTWKGDGMYKKKVKRDTTCTYUDVMAADSWMYTAMMRCDVTTYRNN	577
Qy	405	TTTCCTTTGACATCAAAATTAATTTGTAGACGAGTGTCCATATATAGATGTAAGTAAA	464
Db	576	TYCKSYAHSYMYNNNAAMYRYSARBNSSMARMTTRNNMMMSGBVRMRYAGTWMEHNN	517
Qy	465	TGAATTTAGAGTAAATAAATATATPAAGAAAGTGAACAAAAAATTTAGTGAAAAAGTAA	524
Db	516	NNNTDTRYYWWWRMARBTTTYUDDSCAKSXMRGNRRNARAKMWWAANNDGAAHDHTYW	457
Qy	525	CTGGAAGAAAAGGCTTATTTGGCTTGGCGCAAGTGCATPAAGTTATTTTACCAAAAAGGA	584
Db	456	MGNNTYMMRRAMKMMMMACRAIYCCNNNNNPRACYNHGHKMYRTYKTMWGAACNNNBK	397
Qy	585	AAGAGTACTGGCTTGCTCTCTCTCTTTTAAATGTAACSTGTAAAGATAGAGAGA	644
Db	396	AMYMRVAMMYVSHDTNTDMMNMWTSDBWHTVVDYTWMRAMNNNNNNNNNMBCKTSMW	337
Qy	645	CTCAATTTCTAGTCAATCAAAAATATTTTGGGCTATATAT	684
Db	336	WMDHNTHTCTYGNNTGSAIYBMAAASMMWAAAGSNVTTNW	297

```

RESULT 7
US-09-949-016-15270/c
: Sequence 15270, Application US/09949016
: Patent No. 6812339
:
: GENERAL INFORMATION:
:
: APPLICANT: VENTER, J. Craig et al.
:
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
:
: FILE REFERENCE: CLO01307
:
: CURRENT APPLICATION NUMBER: US/09/949, 016
:
: PRIOR FILING DATE: 2000-04-14
:
: PRIOR APPLICATION NUMBER: 60/241,755
:
: PRIOR FILING DATE: 2000-10-20
:
: PRIOR APPLICATION NUMBER: 60/237,768
:
: PRIOR FILING DATE: 2000-10-03
:
: PRIOR APPLICATION NUMBER: 60/231,498
:
: PRIOR FILING DATE: 2000-09-08
:
: NUMBER OF SEQ ID NOS: 207012

```

```

: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 15270
: LENGTH: 192302
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: (1) .. (192302)
: OTHER INFORMATION: n = A,T,C or G
: US-09-949-016-15270

```

Query Match	7.14	Score	49.8	DB	4	Length	192303
Best Local Similarity	47.53	Pred. No.	0	42			
Matches	184	Conservative	0	Mismatches	197	Indels	6
						Gaps	1

QY	2	TAAATATAATGATTAATAATGGCAATATAAAGTAGATGAGGTTTTTTAAATTTTTTA	83
Db	76211	TGATATATTTGGTTAAAATATCAGAGGTATTTAAATATAAATATAGATGCAATTTAACT	76152
QY	84	TTATAAATTAATTTTTTGCAATGAAATGTTTTTAAGTAAAAATTTTGACATTAACAAGTA	143
Db	76151	TTTTAAAAATCTTTTAAACAAGATACCTTTTTCAGTAAAAAATGTAGACAGTGAATAGA	76092
QY	144	TTATTTTTCCGCAAGATTGAGTCTACGAGAATGTCTT-----GTCTTTAACTCGT	197
Db	76091	GAATTCACCCCTTTTTTGGAGGAACAACGTGTCCCAACTCTAGTAATTTAAACTGTG	76032
QY	198	AAAGTTTTGGTCTTACCACAACCAATATCCACAAAGGTAAAGACCATACCGTTAATTA	257
Db	76031	TTTTTTTTTTTTTTTTTAAAGAAAGTAGCAAGCTGATGAAAAATATCTCTCCGAG	75972
QY	258	TATCTCAACCGGTTTAAGATTACATTAATCATTTACTAATTCGGGTAAATTAAGTTT	317
Db	75971	TACTCTCTGCTGTACTATGTATATGATATGATCAATTTTCATGTCTGCATCACCAATAT	75912
QY	318	TATGAATGTGGTTATTTTGTAGGTACGTGAATTTATTAATTTTTTGTACTGTGTTT	377
Db	75911	TCTAAAGAAGGTCTAAGAAATTTGAGGATTAATGATTTAGTCAATGTTTGTAGTAAATTTT	75852
QY	378	TCTTTTTGGTCAACTAGTACTTT	404
Db	75851	TTGTTTCAAAATTTAATCGAGACATGT	75825

RESULT 8
 US-08-998-416-1137
 Sequence 1137, Application US/08998416
 Patent No. 6239264
 GENERAL INFORMATION:
 APPLICANT: Philippesen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jürgen
 APPLICANT: Knechtle, Philipp
 APPLICANT: Redischung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYPHILII
 TITLE OF INVENTION: AND USBS THEREOF
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416

Qy	61	GGAGAGTTTTTTAAATTTTTTAAATAAATAAATTTTTTGAAGAAATGTTTTTAA	120
Db	61	TGAAGTTTTTTTTAAATTTTTATTAATAAATTTTTTGCATGAATGTTTTTAA	120
Qy	121	TAAATTTTGA CATTAACCAAGATATTTTCGCCACCAATTTAGCTTACGAGATGTC	180
Db	121	TAAATTTTGA CATTAACCAAGATATTTTCGCCACCAATTTAGCTTACGAGATGTC	180
Qy	181	CTGTCTTAAACCTCGTAAAGTTTGGTCTTACCAACCCAAATACCCCAAAAGTTAACG	240
Db	181	CTGTCTTAAACCTCGTAAAGTTTGGTCTTACCAACCCAAATACCCCAAAAGTTAACG	240
Qy	241	ACCAATACCGGTTAATATATTTCTAAACCGGTTATATAGTTACATAAATCATTTACTAATC	300
Db	241	ACCAATACCGGTTAATATATTTCTAAACCGGTTATATAGTTACATAAATCATTTACTAATC	300
Qy	301	CGCGGTATTAATTAAGTTTATGAAATGTGTTATTTTGTAGTCACTGTAATTTTAAAT	360
Db	301	CGCGGTATTAATTAAGTTTATGAAATGTGTTATTTTGTAGTCACTGTAATTTTAAAT	360
Qy	361	TTTTTAGTACTTGTTTTTCTTTTTGGGTCAACTAGTTACTTTTTCTTTTGACATCA	420
Db	361	TTTTTAGTACTTGTTTTTCTTTTTGGGTCAACTAGTTACTTTTTCTTTTGACATCA	420
Qy	421	AATATTGTAGACGAGTGTCCATATATAGATGTGAATGAAATGAATATTTGATTAATA	480
Db	421	AATATTGTAGACGAGTGTCCATATATAGATGTGAATGAAATGAATATTTGATTAATA	480
Qy	481	AATAAATATAGAAAGTGAAACAAAAAAATTAAGTGAAGAAAGTTAAGTGAAGAAAGGCT	540
Db	481	AATAAATATAGAAAGTGAAACAAAAAAATTAAGTGAAGAAAGTTAAGTGAAGAAAGGCT	540
Qy	541	TATTGGCTTGGCAAGATTCATTAAGTTATTTTCAACAAAAAGGAAGAGTCTTGCGCCT	600
Db	541	TATTGGCTTGGCAAGATTCATTAAGTTATTTTCAACAAAAAGGAAGAGTCTTGCGCCT	600
Qy	601	GCTCTCTCTCTTTTAAATGTTAACTGTAAAGTGAAGATAGAGACTCAATTTTCTAGTTCA	660
Db	601	GCTCTCTCTCTTTTAAATGTTAACTGTAAAGTGAAGATAGAGACTCAATTTTCTAGTTCA	660
Qy	661	TCAAAACATATTTGGCGTTATATTTCTGTGTCACTTGAATACC 705	
Db	661	TCAAAACATATTTGGCGTTATATTTCTGTGTCACTTGAATACC 705	

```

RESULT 2
US-10-810-788A-3
; Sequence 3, Application US/10810788A
; Publication No. US20040216184A1
; GENERAL INFORMATION:
; APPLICANT: Ruezinsky, Diane
; TITLE OF INVENTION: Novel Plant Promoters for Use in Early Seed Development
; FILE REFERENCE: REN-00-118
; CURRENT APPLICATION NUMBER: US/10/810,788A
; CURRENT FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: US 60/458,828
; PRIOR FILING DATE: 2003-03-28
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: SeqIdIn version 3.1
; SEQ ID NO 3
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-10-810-788A-3

```

Query Match	99.9%	Score 704;	DB 21;	Length 2017;
Best Local Similarity	100.0%	Pred. No. 7.1e-122;		
Matches 704;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 AAGCTTTAAATTGGGAAAAATTTAATAATATATGCTATTAAATGCCAAATAAAAGTTAGA 60
Db 1314 AAGCTTTAAATTGGGAAAAATTTAATAATATATGCTATTAAATGCCAAATAAAAGTTAGA 1373

QY	61	TGAGAGTTTTTAAATTTTATTTATTTAAATTAATTTTGCATGAATTTGTTTTAAAG	120
Db	1374	TGAGAGTTTTTAAATTTTATTTATTTAAATTAATTTTGCATGAATTTGTTTTAAAG	1433
QY	121	TAAATTTTGACATTAACCAAGTATTATTTCCGCCAGAAATTGATCTACGAGAGATGTC	180
Db	1434	TAAATTTTGACATTAACCAAGTATTATTTCCGCCAGAAATTGATCTACGAGAGATGTC	1493
QY	181	CTGTCTTTAAACCTCCGTAAGTTTTGTCTTACCACCAATACCACAAAGGTAAAG	240
Db	1494	CTGTCTTTAAACCTCCGTAAGTTTTGTCTTACCACCAATACCACAAAGGTAAAG	1553
QY	241	ACCAATACCGGTATATAATATTCTAACCGGTTTAAAGTTTACATAATCATTTACTAATC	300
Db	1554	ACCAATACCGGTATATAATATTCTAACCGGTTTAAAGTTTACATAATCATTTACTAATC	1613
QY	301	CGCGTGTAATTAAGTTTATGAAGATGCGTTATTTTGTAGTGCACTGAAATTTATTAAAT	360
Db	1614	CGCGTGTAATTAAGTTTATGAAGATGCGTTATTTTGTAGTGCACTGAAATTTATTAAAT	1673
QY	361	TTTTTAGTACTGTGTTTTCTTTTGGGTCAACTGATTACTTTTCCTTTGACATCA	420
Db	1674	TTTTTAGTACTGTGTTTTCTTTTGGGTCAACTGATTACTTTTCCTTTGACATCA	1733
QY	421	AATTTATTTGACGAGTGTCCATATATAGATGTGAATGGAATGAATATTGAGTAAAT	480
Db	1734	AATTTATTTGACGAGTGTCCATATATAGATGTGAATGGAATGAATATTGAGTAAAT	1793
QY	481	AATTAATATATGAAGTGAACAAAAAAAATTAAGTGAAAAAGTAACTGGAAGAAAAAGGCT	540
Db	1794	AATTAATATATGAAGTGAACAAAAAAAATTAAGTGAAAAAGTAACTGGAAGAAAAAGGCT	1853
QY	541	TATTGGCTTGCAAGATTCCATAAGTTTATTTACCAAAAAAGGAAGAATCTTGCGCTT	600
Db	1854	TATTGGCTTGCAAGATTCCATAAGTTTATTTACCAAAAAAGGAAGAATCTTGCGCTT	1913
QY	601	GCTCTCTCTCTTTTAAATGTTAAACGTGTAAGATGAGAGACTCAATTTCTAGTTCA	660
Db	1914	GCTCTCTCTCTTTTAAATGTTAAACGTGTAAGATGAGAGACTCAATTTCTAGTTCA	1973
QY	661	TCAAAAATATTTTGGCGTTATTTCTGTGGTCACTGTAATC	704
Db	1974	TCAAAAATATTTTGGCGTTATTTCTGTGGTCACTGTAATC	2017

```

RESULT 3
US-10-473-126-386/c
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; PILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

```

Query Match	9.1%	Score 64.2;	DB 21;	Length 8056;
Best Local Similarity	45.7%;	Pred. No. 0.059;		
Matches 260;	Conservative 0;	Mismatches 308;	Indels 1;	Gaps 1;

QY 5 TTTTAAATTTGGGAAATTTTAAATATATGATTTAAATGCGAAATAAAGTAGATGAC 64

Db 7154 TTACCAAAATTATATATATATTTTAAATTTTAAAAAAAATTATTAATTAAATAAAT 70395

Query Match	8.9%	Score 62.6	DB 21	Length 8056
Best Local Similarity	45.0%	Pred. NO. 0.12		
Matches 274	Conservative	0	Mismatches 334	Indels 1
			Gaps	1

[illegible]

```

OY      191 ACCCTGTAAGGTTGGTTCCTTACCCCAAGCCAAATACCACAAAGGTAAACGACCATATCCGG 250
Db      2289 AATTATTTATTTATTTTAAAATTAATTATTTAATTTTGTATTTTAATTAATTAATTTTATTTAATT 2348
OY      251 TTAATTAATATTTCTPAACCGGTTTATTAAGTTTACATAAATCATTTACTAATCCGGCTTAAT 310
Db      2349 TTTAAATATTAATTTTATTTTATTTATTAATAATTAATTTAATTTATTTATTTATTTTGAAAAAT 2408
OY      311 TTAAGTTTATGAAAGTGTGTTATTTTGTAGGTGACGTGAAATTTATTAATTTTGTAGTAC 370
Db      2409 TAAAAAATAATTAATTTTATTTATTTTAAAAATTAATTTTAAAAATTTTAAATTTTAAATTTTA -TAT 2467
OY      371 TTGTTTTCTTTTGGGTTCACTAGTACTTTTTCTTTGACATCAAAAATTTATTTGA 430
Db      2468 ATTTTTTATTTATTTTGTTTTAAAAATTAATTTAATTTAATGAAAAATTAATAAAAAATTTAAA 2527
OY      431 GACGAGTGTCTCATATTAATAGATGTGAAATGAATGAATATTTAGTAATTAATTAATATA 490
Db      2528 TTTAAATTTAAAAATTAATTTAAAAAATTTAAAAATTTAAATTTAAAAATTTAAAAATTA 2587
OY      491 GAAATGAAACAAAAAATTTAGTGGAAAAAGTACTGGAAGAAAAAGGCTATATTGCTTG 550
Db      2588 AATAAATGAAAAAATAATTAATTAATTAATTAATTTAAAAAATAAAAAAATAAAAAAATAAT 2647
OY      551 GCAAGATTCCATTAAGTTTATTTTACCAAAAAAGAAAGAGTACTGTGGCTGTCTCTC 610
Db      2648 GTTAAAAATTTAAAAATTTTAAAAATTAATAATTTTAAAAATTTTAAATTTTAAATTTATTTT 2707
OY      611 TCTTTTAAA 619
Db      2708 TTTTTTTAA 2716

RESULT 5
US-10-021-323-6375
; Sequence 6375, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Peng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021.323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17980
; SEQ ID NO 6375
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(531)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3828-004-Q1-N6-H2
; US-10-021-323-6375

```

[illegible]


```
RESULT 8
US-10-257-166-17
; Sequence 17, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 17
; LENGTH: 11422
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-257-166-17

Query Match      8.6%; Score 60.6; DB 18; Length 11422;
Best Local Similarity 48.0%; Pred. No. 0.31;
Matches 272; Conservative 0; Mismatches 284; Indels 11; Gaps 3;

QY 6 TTTAAATTTGGGAAATTTAATAATATATATATTAATGCAATTAAGTTAGAGAGA 65
DB 7207 TTTAAATTTGTTAGGAAATTTAATATATATTTTAAATTTGAAATTAATAAATA 7266
QY 66 GTTTTAAATTTTATTTATTAATAATATTTTTCATGAAATTTGTTTAAAGATAAA 125
DB 7267 GATTCTTATTTTATTTATTTGAAAGTATTTGTTAATAATGTTAGATTTT-----TTAAA 7322
QY 126 TTTTGACATPACCAAGATTTATTTTCCGCCAGATGAGTCTACAGAGAGTCTGTC 185
DB 7323 AATTGTTAATTTATTTTATTTTAAATTTAAATTTAGTTTAAATTTAAATTT 7382
QY 186 TTTTAAATTTTAAATTTTGTCTTACCAACCAATCCCAAGGTAAGCAACAT 245
DB 7383 TTTTAAATTTTAAATTTTAAATTTGAAATGTTGAAATTTTAAATTTTAAATTTT 7438
QY 246 ACCGGTAAATATTTCTAACCGGTTTATTAAGTTTACATPAAATCTTACTATCCGGCT 305
DB 7439 TGTTTTATTTTATTTAATTTAATTTATTTTGTAAATGTAATTTAATTTAATTTAGATAT 7498
QY 306 GTAATTAAGTTTATGAATTTGTTATTTTGTAGTCAAGTAAATTTTATTAATTTT 365
DB 7499 ATAGTGTATTTGTAATAATTTGTTTATTTTATATATATGTTTATGTTATATGTTT 7558
QY 366 AGTACTTGTCTTTTCTTTTGGGTTCACTAGTACTTTTCTTCTTGACATPAAATTA 425
DB 7559 TTTAGATTAATGAGATGATTTATTTATTTGATTAATTTTAAATTTTAAATTTTAAATTT 7615
QY 426 TTGTAGACAGTGTGTCATATATAGATGGAATGAATGAATATTTGATTAATAATA 485
DB 7616 TTTAATTTTACGTTGTGTATATGTTTAAATTTTAAAGTATTAATTAATAATAATA 7675
QY 486 ATATAGAAAGTGAACAAAAAATTAAGTGAAGAAAGTGAAGAAAGAAAGGCTTATG 545
DB 7676 AAAAAAAGAAAAAAGCAATATGTTTGTGAATTTTAAATTTAATAATAATAATTTTGAAG 7735
QY 546 GCTTGCAAGATTCCTAAGTTATTT 572
DB 7736 TATAGATTAAATATTTTATTTATTT 7762
```

```
RESULT 9
US-10-021-323-2253/C
; Sequence 2253, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd B.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21 (52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 2253
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Geosyrium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3825-011-01-N6-C9
US-10-021-323-2253

Query Match      8.6%; Score 60.4; DB 20; Length 560;
Best Local Similarity 50.7%; Pred. No. 0.12;
Matches 145; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 251 TTAATATATTTCTAAACCGGTTTATTAAGTTTACATTAATCATTTACTAATCCGGTAAAT 310
DB 371 TTTTATTTTATTTTAAACCACTTTTATTTTAAATTTTATTTTATTTTATTTTATTTT 312
QY 311 TAAATTTAAGAAATGTTGTTATTTTGTAGTCAGTGAATTTATTAATTTTATTTAGTAC 370
DB 311 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 252
QY 371 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 430
DB 251 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 192
QY 431 GACGAGTGTCCATATATATGATGTCGAAATGAATGATATTTGATTAATAATAATA 490
DB 191 AATATATTTTATTTTAAATTAATAATAATAATAATAATAATAATAATAATAATAATA 132
QY 491 GAAAGTGAACAAAAAATTTAGTGAAGAAAGTAACTGAAAGAAA 536
DB 131 AAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 86

RESULT 10
US-10-357-930-60960/C
; Sequence 60960, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Emdege, Wilson
; APPLICANT: Monahan, John
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
; TITLE OF INVENTION: HUMAN PROSTATE CANCER
; FILE REFERENCE: MRI-007BCN
; CURRENT APPLICATION NUMBER: US/10/357, 930
; CURRENT FILING DATE: 2003-02-04
; PRIOR APPLICATION NUMBER: 09/785, 276
; PRIOR FILING DATE: 2003-02-16
; PRIOR APPLICATION NUMBER: 60/183, 319
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 60/189, 862
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/207, 454
; PRIOR FILING DATE: 2000-05-25
```

```
; PRIOR APPLICATION NUMBER: 60/211,314
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/219,007
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/255,281
; PRIOR FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 62232
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60960
; LENGTH: 626
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 48, 49, 50, 117, 135, 136, 137, 166, 168, 189, 194, 197,
; LOCATION: 278, 281, 293, 297, 281, 291, 323, 324, 327, 328,
; LOCATION: 329, 331, 336, 337, 338, 339, 371, 372, 376, 377, 379, 381,
; LOCATION: 382, 394, 403, 404, 405, 406, 408, 413, 429, 431, 437
; OTHER INFORMATION: n = A,T,C or G
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 447, 451, 453, 455, 456, 461, 469, 484, 487, 493, 499, 509,
; LOCATION: 515, 529, 538, 540, 541, 542, 545, 558, 559, 561, 564, 565,
; LOCATION: 581, 582, 584, 585, 589, 592, 598, 599, 601, 606, 609, 617,
; LOCATION: 618
; OTHER INFORMATION: n = A,T,C or G
;
; US-10-357-930-60960

Query Match      8.5%; Score 60.2; DB 21; Length 626;
Best Local Similarity 48.0%; Pred. No. 0.14;
Matches 134; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 258 TATTCGTAACCGGTTATATAGTTTACATTAATCATTTACTAATCCGGGTAAATTAAGTT 317
DB 330 TNNNCNCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 271
QY 318 TATGAAATGCGTATATTTGTAGTCACGTGAAATTTATTAATTTTGTAGTCTGTTT 377
DB 270 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 211
QY 378 TCTTTTGGGTTCACAGTACTGTTCTTTCTTTGACATCAAAATATTTAGTACTGTTT 437
DB 210 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 151
QY 438 GGTCCATATATATGATGTAATGAAATGAAATGATTTAGTAAATTAATATATGAAGTG 497
DB 150 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 91
QY 498 AACCAAAAAAATTAAGTGAAGAAAGTAACTGAAAGAAAA 536
DB 90 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 52

RESULT 11
US-10-021-323-1716
; Sequence 1716, Application US/10021323
; Publication No. US2004012340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, J111
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd B.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 1716
; LENGTH: 499
; TYPE: DNA
```

```
; ORGANISM: Gossypium hirsutum
;
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB825-005-Q1-N6-D10
; US-10-021-323-1716

Query Match      8.5%; Score 60; DB 20; Length 499;
Best Local Similarity 50.7%; Pred. No. 0.14;
Matches 144; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 305 TGTAAATTAAGTTTATGAAATGTGTTATTTTGTAGTCACGTGAATTTATTAATTTT 364
DB 34 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 93
QY 365 TAGTACTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 424
DB 94 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 153
QY 425 ATTGTAGACGAGTGTCCATATATATGATGTGAATGAATGAATATTTAGTAAATTA 484
DB 154 AAAAAAGGGGGGTAAAAAATTTTAAAAAGAAAAAAGAGGGGGGTAAAAA 213
QY 485 AATATGAAAGTGAACAAAAAATTAAGTGAAGAAAGTAACTGAAAGAAAAAGCTTATT 544
DB 214 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGGGGAAAAA 273
QY 545 GCGTTGCAAGATTTCATTAAGTTTATTTTCAACCAAAAGGAAGA 588
DB 274 AAAATGGGAAAAAAGGGGTTTAAAAAAGGGGTTTAAAAA 317

RESULT 12
US-10-473-126-240
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Biogenics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: Proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-473-126-240

Query Match      8.4%; Score 59.4; DB 21; Length 8056;
Best Local Similarity 44.7%; Pred. No. 0.46;
Matches 272; Conservative 0; Mismatches 336; Indels 1; Gaps 1;

QY 11 ATTGGGAAAAATTTAATAATATATGTAATTAATGCAATTAAGTTAGATGAGTTT 70
DB 2109 ATTATATATATATTTAATATATTTTATTTATTTAATTAATTAATTAATTAATTAATTA 2168
QY 71 TTTAATTTTATTAATAAATTAATTTTTCATGATGAATTTGTTTAAAGATAAATTTTG 130
DB 2169 ATTATATATTTTATTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2228
QY 131 ACATTAACAAGATATATTTTCGCAAGATTAAGTCAAGAGATGTCGTCTTAA 190
DB 2229 AATTTTAATTAATAATTTATTTTATTAATTAATTAATTAATTAATTAATTAATTTT 2288
QY 191 ACCTGTAAGTTTGTGTTTACCAACCAATTAACCAAAAGTAAAGACATACCGG 250
DB 2289 AATTTATTTATTTTAAATTAATTTATTAATTTTGTATTTAAATTAATTTATTTATTT 2348
QY 251 TTAATATATTCGACCGGTTATTAAGTTTACATTAATCATTTACTAATCCGCTGAAT 310
DB 2349 TTTAATATATATTTTATTTTATTAATTAATTTATTAATTTTATTTTATTTTTCGAAAT 2408
```



```

? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (6666) .. (66660)
? OTHER INFORMATION: n is a, c, g, or t
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (6663) .. (6663)
? OTHER INFORMATION: n is a, c, g, or t
? FEATURE:

```

Oy	5	TTTTAAATTTGGGAAAATTTAATAATTAATGAATTAATAGCAAGAATGAAGCTACATGAC	64
Dd	6653	TTTATATANNNAANTPAAAAAATATATATANNAAAAAATATATATNNAAAAAAT	6594
Oy	65	AGTTTTTTAAATTTTATATATAAATAATTTTCATGAATAATGTCTTTTAAAGATAAA	124
Dd	6593	TTTTTTATFAAATTTTTTAAAAAAAAANNTTTTTTTTTAAAAAATATTTATATANN	6534
Oy	125	ATTTTGACATPACCAGTATTAATTTCCGCACGAATGAGTCAGAGAGATGCTGT	184
Dd	6533	TTTTTATTTTAAATTTNTTTATTTTAAAAATAAAAATATNTAAAAAANNTATTTTTTT	6474
Oy	185	CTTTAAACCTCGTAAAGTTTGGCTTACCCAACCCAATACCCCAAAAGGTAAACGACA	244
Dd	6473	AAAAAANNTNAAAAATATTTTTTTTTATTTTT-----TTNCGNAAAAAATAATAATPA	6419
Oy	245	TACCGGTTAATAATATCTAAACCGGTTATPATAGTTTACATPAATCATTTACTAATCCCGC	304
Dd	6418	NTNANNTPAAAAATTTTTTTTTTAAATAAANTNTTTATATAAAAAATATTTTTT----	6364
Oy	305	TGTATATTAAGTTTATGAATATGTCGTATTTTGAAGCACGGAATTTTATATTTTT	364
Dd	6363	--TAAAAAANAATAAATTTNNNTTTTTTTTATATNNNNNTTATTTATTTTTTTNT	6306
Oy	365	TAGTACTGTCTTTCTTTTGGGCTCAACTAGTACTCTTTTCTTGTGACATCAAAAT	424
Dd	6305	NTTTTNNNNNNNNNTTTTTTTTTNNNNNTTTTTTTTTTTTTNTNNNNNTTTNTWT	6246

Query Match	8.4%	Score 59.2;	DB 22;	length 50000;
Best Local Similarity	54.8%	Pred. No. 0.94;		
Matches 138; Conservative	0;	Mismatches 113;	Indels 1;	Gaps 1;

QY 260 TTCTAACCGTTTAAAGTTTACATAAATCATTTACTAATCCGGTGAATTAAGTTT 319
DB 49095 TTATTACTTATTTTATTTATTTTATTCATCATTTTATTAATTAATTTATTCAAATTA 49154
QY 320 TGAATGTTGTTTATTTTGTAGCGTGAATTTATTAATTTTATTTTGTAGTCTGTTTTC 379
DB 49155 TAAATATTTTATTTTATTTTATTTTAAAGATTTTATTTTATTTTGTATGTTATTAAT 49214
QY 380 TTTTGGGTTCACTAGTACTTTTCTTGTACATCAAAATTAATTTAGTGAAGTGG 439
DB 49215 AAATTTATTAATTAATTAATTAATTTATTTTAAATCAATTAATTTTAAATGCGA 49274
QY 440 TCCATATATAGATGTGAATGAATGAATATTTGAGTAATTAATTAATTAATGAAGTGA 499
DB 49275 AAATTAATTAATGAATTAATTTATTTATTTA -GATTTTCAATTAATTAATTAATTTGTT 49333
QY 500 CAAAAAAATTA 511
DB 49334 CTAAATTAATTA 49345

RESULT 15

US-10-021-323-16830
; Sequence 16830, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 16830
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3829-024-Q6-N6-E4
US-10-021-323-16830

Query Match 8.4%; Score 59; DB 20; Length 469;

Best Local Similarity 51.3%; Pred. No. 0.21;

Matches 137; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 270 TTATAAGTTTACATAAATCATTTACTAATCCGGTGAATTAAGTTTATGAATGTGG 329
DB 11 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 70
QY 330 TTATTTTGTAGTCACGTGAATTTATTAATTTTATTTTGTAGTCTGTTTCTTTTGGCT 389
DB 71 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 130
QY 390 TCAACTAGTACTTTTCTTGTGACATCAAAATTAATTTGAGCGAGTGTCCATATATA 449
DB 131 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 190
QY 450 GATGGAATGAATGAATTAATTTAGTATTAATTAATTAATTAATTAATTAATTAATTAAT 509
DB 191 AAAAAAAAAAAAAAAAAAAAAAAAAATTTTATTAATTAATTAATTAATTAATTAATTAATTA 250
QY 510 TAGTGAAAAAGTAACTGAAAAAGAAA 536
DB 251 TTTAATAAAAAAAAAAAAAAAAAAAAAA 277

Search completed: November 17, 2005, 18:03:54
Job time : 855 secs

This Page Blank (uspto)